Poster Evolution from 2014-2021

ML Tlachac

Why am I Presenting About Posters?

18 Posters Created

12 Presentations

8 Venues

8 Projects

Stage 1: Undergraduate

Pattern Avoidance on Forests of Binary Shrubs

David Bevan[◁], Derek Levin*, Peter Nugent*, Jay Pantone[⊕], Lara Pudwell[†], Manda Riehl*, ML Tlachac*

*University of Wisconsin-Eau Claire, †Valparaiso University, ◁Open University London, and ⊕University of Florida

1. Introduction

We examine labeled increasing forests composed of binary trees with three nodes whose associated permutations avoid patterns of length 3 and investigate how these forests are related to other combinatorial objects. This work continues a long line of research extending the notion of classical pattern avoidance on permutations to other structures, for example: Dyck paths (West and others), binary trees (Pudwell and others), linear extensions of posets (Yakoubov and others), heaps (Riehl and others), and polyominoes (Yan and others).

The most general form of the question considered here has nice applications to periodic scheduling problems, where there are recurring precedence requirements. For example, consider a scenario with n identical robots each executing a task, such as retrieving a box and adding it to a stack. In such a scenario, we have not only precedence requirements on the actions of each individual robot (the increasing tree requirement) but also global precedence requirements. For example, ensuring that no two heavier boxes are both stacked on top of a lighter box is equivalent to avoiding the patterns 231 and 321.

2. Definitions

Given permutations $\pi = \pi_1\pi_2\cdots\pi_n$ and $\rho = \rho_1\rho_2\cdots\rho_m$ we say that π **contains** ρ as a pattern if there exist $1 \le i_1 < i_2 < \cdots i_m \le n$ such that $\pi_{i_a} < \pi_{i_b}$ if and only if $\rho_a < \rho_b$. In this case we say that $\pi_{i_1}\pi_{i_2}\cdots\pi_{i_m}$ is **order – isomorphic** to ρ and that $\pi_{i_1}\pi_{i_2}\cdots\pi_{i_m}$ **reduces to** ρ. If π does not contain ρ, then π is said to **avoid** ρ.

An **increasing tree** is a tree where each child node has a label larger than the label of its parent node. Within these trees and forests is the **associated permutation**, which is the permutation formed by reading vertex labels from left to right across each level of the tree starting at the root. Next, a **binary tree** is a tree where each node has 0 or 2 children.

3. Examples

The permutation 2 4 1 3 contains an occurrence of 132, because the sub-permutation 2 4 3 reduces to 1 3 2. However, 2 4 1 3 avoids the patterns 123 and 321.

Below we have a forest of increasing binary trees:

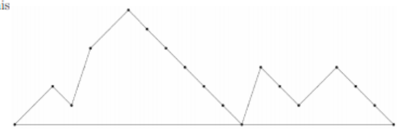
Associated permutation: 10 12 11 1 2 9 3 4 8 5 6 7

Contains patterns: 123, 132, 231, 312, and 321

Avoids pattern: 213

4. Bijection between forests avoiding 213 and lattice paths from $(0,0)$ to $(4n,0)$ using steps $(1,3),(2,2),(1,-1)$ staying weakly above the x-axis

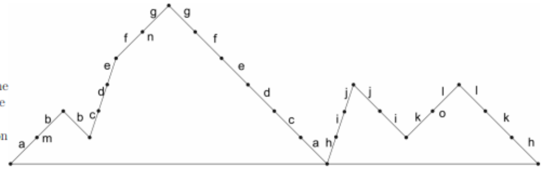
Start with a lattice path from $(0,0)$ to $(4n,0)$ using steps $(1,3),(2,2),(1,-1)$ staying weakly above the x-axis. This path is from (0,0) to (20, 0), corresponding to $n = 5$ trees.

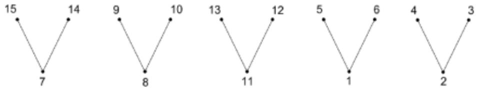


Walk along the path from left to right, labeling each vertical increase alphabetically. We use letters a through l for this example path. Next, at the middle of each $(2,2)$ step, label the middle point with a unique letter. This path has three $(2,2)$ steps, so we label the middle points m, n, and o.

Label each $(1,-1)$ downstep with the corresponding vertical increase label. Thus, we will have downsteps with labels a through l.

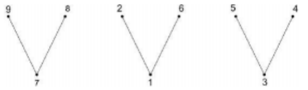
Read from left to right and record the first occurrence of each letter. In doing this we obtain the word $a\ m\ b\ c\ d\ e\ f\ n\ g\ h\ i\ j\ k\ o\ l$. Then reading from right to left on the path, number the letters in the above string in the order in which they appear. The result is the permutation associated with the lattice path. In this instance, it is 7 15 14 8 9 10 11 13 12 1 5 6 2 4 3. Here is the permutation displayed on a binary forest. This is a demonstration of how to use the bijection to take a lattice path to a forest.

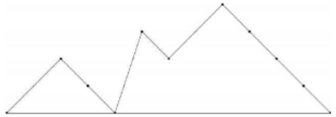




Now, let us reverse the bijection and find the lattice path corresponding to a given binary forest. For this example, we use the permutation 7 9 8 1 2 6 3 5 4 which is read from the forest of 3 trees from below.

In the permutation, mark all of the numbers that start a decrease between consecutive leaves. In our permutation, the marked numbers are 9 and 5. Now we label the remaining numbers alphabetically from left to right. This results in a 9 $b\ c\ d\ e\ f$ 5 g. Then we label the the marked numbers with their own unique marked alphabetical letters, in this instance let us choose h and i. Thus our permutation is now $a\ h\ b\ c\ d\ e\ f\ i\ g$, where h and i are considered marked.





Now we build the path. Each increase between consecutive leaves corresponds to a $(1,3)$ increasing step and each decrease between consecutive leaves corresponds to $(2,2)$ increasing step. Then to find the number of $(1,-1)$ decreasing steps between each increasing step, we look at the root of the next tree and count the number of unmarked letters prior to it which are larger. Next, we subtract the number of decreasing steps we have built prior to these. Then continue these steps until there is a corresponding increase for every tree and that the end point is at $(12,0)$. The completed path for our example is shown above.

5. Table

Increasing forests of shrubs	Sequence	OEIS#
Total	$2, 80, 13440, 5913600, \ldots$	$A210277 : (3n)!/3^n$
$Av(123)$	$1, 3, 12, 55, 273, \ldots$	$A001764$
$Av(132)$	$1, 4, 22, 140, 969, \ldots$	$A002293$
$Av(213) = Av(312)$	$2, 14, 134, 1482, 17818, \ldots$	$A144097$
$Av(231)$	$2, 23, 377, 7229, 151491, \ldots$	$A060941$
$Av(321)$	$1, 37, 866, 23285, 679606, \ldots$	new
$Av(132, 213) = Av(132, 312)$	$1, 2, 4, 8, 16, \ldots$	$A000079 : 2^k$
$Av(132, 321)$	$1, 4, 10, 19, 31, \ldots$	$A005448$
$Av(213, 231) = Av(231, 312)$	$2, 8, 32, 128, 512, \ldots$	$2 * A000302 : 2 * 4^k$
$Av(213, 312) = Av(213, 231, 312) = Av(213, 231, 312, 321)$	$2, 2, 2, 2, 2, \ldots$	$A007395$: continuous 2s
$Av(213, 321)$	$2, 6, 13, 23, 36, \ldots$	$A143689$
$Av(231, 321)$	$2, 12, 72, 432, 2592, \ldots$	$2 * A000400 : 2 * 6^k$
$Av(312, 321)$	$2, 10, 50, 250, 1250, \ldots$	$2 * A000351 : 2 * 5^k$
$Av(132, 213, 321)$	$1, 2, 3, 4, 5, \ldots$	$A000027$: natural numbers
$Av(213, 231, 321)$	$2, 4, 6, 8, 10, \ldots$	$2 * A000049 : 2 * 2^k$
$Av(231, 312, 321)$	$2, 6, 18, 54, 162, \ldots$	$2 * A000244 : 2 * 3^k$

6. Conclusions, conjectures, and future research

In addition to the bijection shown in the poster, we have also found interesting bijections for many other sets of pattern avoiding forests such as in the table above. For forests avoiding 312, we have a bijection to the same lattice paths as the forests avoiding 213 shown, but the paths are read and formed in a slightly different manner. Additionally we have a bijection from forests avoiding 231 to lattice paths of north and east steps weakly below the line $2x/3$. We also have bijections for forests avoiding 123 and 132 to different sets of lattice paths.

Currently, we are working on forests avoiding 321. We have found the quantity of forests which avoid this pattern for forest sizes from 1 tree to 250 trees using a clever recurrence relation; however, it does not coincide with any known sequences.

Although we have made additional bijections to forests avoiding two or more patterns as well, there are some that we are still working on. However, by going through all of these examples, we have decided that our next step has two potential directions. One is to extend our results to k-ary shrubs, which are forests of trees that are a maximum of one level high and contain $k-1$ leaves each. The second option would be to extend our results to forests of binary trees with multiple levels. We will next investigate the k-ary shrubs before moving on to the other option because we know from computations that the shrubs will have fewer unique patterns.

References

[1] David Callan, *Some Bijections for Restricted Motzkin Paths* arXiv:math/0407374v1 [math.CO] (2004).

[2] Derek Levin, Lara Pudwell, Manda Riehl, Andrew Sandberg, *Pattern Avoidance on Heaps*. Accepted to Australasian Journal of Combinatorics.

[3] OEIS Foundation Inc., The On-Line Encyclopedia of Integer Sequences, http://oeis.org.

Acknowledgments

• Office of Research and Sponsored Programs, UW-Eau Claire
• Student Blugold Commitment Differential Tuition funds through the University of Wisconsin-Eau Claire
• UW- Eau Claire Department of Mathematics

Year: 2015
Tool: Latex

Worcester Polytechnic Institute

Pattern Avoidance on Forests of Binary Shrubs

David Bevan[◁], Derek Levin[*], Peter Nugent[*], Jay Pantone[⊕], Lara Pudwell[†], Manda Riehl[*], ML Tlachac[*]

[*]University of Wisconsin-Eau Claire, [†]Valparaiso University, [◁]Open University London, and [⊕]University of Florida

1. Introduction

We examine labeled increasing forests composed of binary trees with three nodes whose associated permutations avoid patterns of length 3 and investigate how these forests are related to other combinatorial objects. This work continues a long line of research extending the notion of classical pattern avoidance on permutations to other structures, for example: Dyck paths (West and others), binary trees (Pudwell and others), linear extensions of posets (Yakoubov and others), heaps (Riehl and others), and polyominoes (Yan and others).

The most general form of the question considered here has nice applications to periodic scheduling problems, where there are recurring precedence requirements. For example, consider a scenario with n identical robots each executing a task, such as retrieving a box and adding it to a stack. In such a scenario, we have not only precedence requirements on the actions of each individual robot (the increasing tree requirement) but also global precedence requirements. For example, ensuring that no two heavier boxes are both stacked on top of a lighter box is equivalent to avoiding the patterns 231 and 321.

2. Definitions

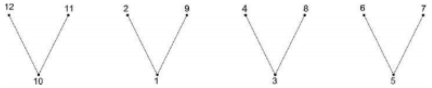
Given permutations $\pi = \pi_1\pi_2\cdots\pi_n$ and $\rho = \rho_1\rho_2\cdots\rho_m$ we say that π **contains** ρ as a pattern if there exist $1 \leq i_1 < i_2 < \cdots i_m \leq n$ such that $\pi_{i_a} < \pi_{i_b}$ if and only if $\rho_a < \rho_b$. In this case we say that $\pi_{i_1}\pi_{i_2}\cdots\pi_{i_m}$ is **order – isomorphic** to ρ and that $\pi_{i_1}\pi_{i_2}\cdots\pi_{i_m}$ **reduces to** ρ. If π does not contain ρ, then π is said to **avoid** ρ.

An **increasing tree** is a tree where each child node has a label larger than the label of its parent node. Within these trees and forests is the **associated permutation**, which is the permutation formed by reading vertex labels from left to right across each level of the tree starting at the root. Next, a **binary tree** is a tree where each node has 0 or 2 children.

3. Examples

The permutation 2 4 1 3 contains an occurrence of 132, because the subpermutation 2 4 3 reduces to 1 3 2. However, 2 4 1 3 avoids the patterns 123 and 321.

Below we have a forest of increasing binary trees:



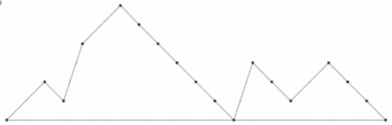
Associated permutation: 10 12 11 1 2 9 3 4 8 5 6 7

Contains patterns: 123, 132, 231, 312, and 321

Avoids pattern: 213

4. Bijection between forests avoiding 213 and lattice paths from $(0,0)$ to $(4n,0)$ using steps $(1,3),(2,2),(1,-1)$ staying weakly above the x-axis

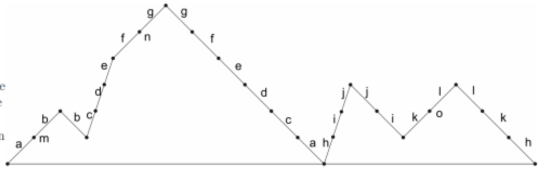
Start with a lattice path from $(0,0)$ to $(4n,0)$ using steps $(1,3),(2,2),(1,-1)$ staying weakly above the x-axis. This path is from (0,0) to (20, 0), corresponding to $n = 5$ trees.

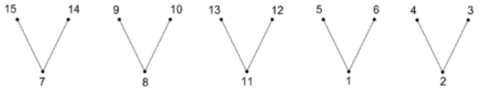


Walk along the path from left to right, labeling each vertical increase alphabetically. We use letters a through l for this example path. Next, at the middle of each $(2,2)$ step, label the middle point with a unique letter. This path has three $(2,2)$ steps, so we label the middle points m, n, and o.

Label each $(1,-1)$ downstep with the corresponding vertical increase label. Thus, we will have downsteps with labels a through l.

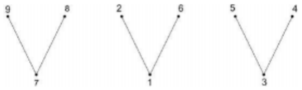
Read from left to right and record the first occurrence of each letter. In doing this we obtain the word $a\ m\ b\ c\ d\ e\ f\ n\ g\ h\ i\ j\ k\ o\ l$. Then reading from right to left on the path, number the letters in the above string in the order in which they appear. The result is the permutation associated with the lattice path. In this instance, it is 7 15 14 8 9 10 11 13 12 1 5 6 2 4 3. Here is the permutation displayed on a binary forest. This is a demonstration of how to use the bijection to take a lattice path to a forest.

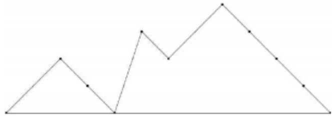




Now, let us reverse the bijection and find the lattice path corresponding to a given binary forest. For this example, we use the permutation 7 9 8 1 2 6 3 5 4 which is read from the forest of 3 trees from below.

In the permutation, mark all of the numbers that start a decrease between consecutive leaves. In our permutation, the marked numbers are 9 and 5. Now we label the remaining numbers alphabetically from left to right. This results in $a\ 9\ b\ c\ d\ e\ f\ 5\ g$. Then we label the the marked numbers with their own unique marked alphabetical letters, in this instance let us choose h and i. Thus our permutation is now $a\ h\ b\ c\ d\ e\ f\ i\ g$, where h and i are considered marked.





Now we build the path. Each increase between consecutive leaves corresponds to a $(1,3)$ increasing step and each decrease between consecutive leaves corresponds to $(2,2)$ increasing step. Then to find the number of $(1,-1)$ decreasing steps between each increasing step, we look at the root of the next tree and count the number of unmarked letters prior to it which are larger. Next, we subtract the number of decreasing steps we have built prior to these. Then continue these steps until there is a corresponding increase for every tree and that the end point is at $(12,0)$. The completed path for our example is shown above.

5. Table

Increasing forests of shrubs	Sequence	OEIS#
Total	$2, 80, 13440, 5913600, \ldots$	$A210277 : (3n)!/3^n$
$Av(123)$	$1, 3, 12, 55, 273, \ldots$	$A001764$
$Av(132)$	$1, 4, 22, 140, 969, \ldots$	$A002293$
$Av(213) = Av(312)$	$2, 14, 134, 1482, 17818, \ldots$	$A144097$
$Av(231)$	$2, 23, 377, 7229, 151491, \ldots$	$A060941$
$Av(321)$	$1, 37, 866, 23285, 679606, \ldots$	new
$Av(132, 213) = Av(132, 312)$	$1, 2, 4, 8, 16, \ldots$	$A000079 : 2^k$
$Av(132, 321)$	$1, 4, 10, 19, 31, \ldots$	$A005448$
$Av(213, 231) = Av(231, 312)$	$2, 8, 32, 128, 512, \ldots$	$2 * A000302 : 2 * 4^k$
$Av(213, 312) = Av(213, 231, 312) = Av(213, 231, 312, 321)$	$2, 2, 2, 2, 2, \ldots$	$A007395$: continuous 2s
$Av(213, 321)$	$2, 6, 13, 23, 36, \ldots$	$A143689$
$Av(231, 321)$	$2, 12, 72, 432, 2592, \ldots$	$2 * A000400 : 2 * 6^k$
$Av(312, 321)$	$2, 10, 50, 250, 1250, \ldots$	$2 * A000351 : 2 * 5^k$
$Av(132, 213, 321)$	$1, 2, 3, 4, 5, \ldots$	$A000027$: natural numbers
$Av(213, 231, 321)$	$2, 4, 6, 8, 10, \ldots$	$2 * A000049 : 2 * 2^k$
$Av(231, 312, 321)$	$2, 6, 18, 54, 162, \ldots$	$2 * A000244 : 2 * 3^k$

6. Conclusions, conjectures, and future research

In addition to the bijection shown in the poster, we have also found interesting bijections for many other sets of pattern avoiding forests such as in the table above. For forests avoiding 312, we have a bijection to the same lattice paths as the forests avoiding 213 shown, but the paths are read and formed in a slightly different manner. Additionally we have a bijection from forests avoiding 231 to lattice paths of north and east steps weakly below the line $2x/3$. We also have bijections for forests avoiding 123 and 132 to different sets of lattice paths.

Currently, we are working on forests avoiding 321. We have found the quantity of forests which avoid this pattern for forest sizes from 1 tree to 250 trees using a clever recurrence relation; however, it does not coincide with any known sequences.

Although we have made additional bijections to forests avoiding two or more patterns as well, there are some that we are still working on. However, by going through all of these examples, we have decided that our next step has two potential directions. One is to extend our results to k-ary shrubs, which are forests of trees that are a maximum of one level high and contain $k-1$ leaves each. The second option would be to extend our results to forests of binary trees with multiple levels. We will next investigate the k-ary shrubs before moving on to the other option because we know from computations that the shrubs will have fewer unique patterns.

References

[1] David Callan, *Some Bijections for Restricted Motzkin Paths* arXiv:math/0407374v1 [math.CO] (2004).

[2] Derek Levin, Lara Pudwell, Manda Riehl, Andrew Sandberg, *Pattern Avoidance on Heaps*. Accepted to Australasian Journal of Combinatorics.

[3] OEIS Foundation Inc., The On-Line Encyclopedia of Integer Sequences, http://oeis.org.

Acknowledgments

- Office of Research and Sponsored Programs, UW-Eau Claire
- Student Blugold Commitment Differential Tuition funds through the University of Wisconsin-Eau Claire
- UW- Eau Claire Department of Mathematics

Year: 2015
Tool: Latex

Self Critiques

- Too many words
- Words too small
- Too much future research
- Appeals to small audience
- Good use of examples with diagrams

Sociodemographic Factors Influencing Household Energy Efficiency in the United States

ML Tlachac

Motivation

Knowing which people are more energy efficient is increasingly important as the negative effects of local pollution, global warming, and energy insecurity are becoming more pronounced. Determining the sociodemographic factors of the households which are more energy efficient allows for better targeting of marketing and policies.

Energy efficiency is reducing the amount of energy needed to perform a task

Data

This analysis incorporates multiple studies: the Residential Energy Consumption Survey (RECs) 2009 and the American Housing Survey (AHS) 2011 are used for the analysis. RECs 2009 surveyed 12083 households and AHS 2011 surveyed 68162 households within the United States. RECs include more questions about energy efficiency while AHS includes more sociodemographic information. This analysis studies multiple aspects of energy efficiency available in both surveys.

There are two different energy efficient factors for each Energy Star appliance: people can actively choose to purchase an Energy Star appliance and they can have an Energy Star appliance. These have very different implications for marketing and policies.

Energy Star appliance availability was unavailable in the literature. As such, a study incorporating 225 refrigerators, 100 dishwashers, and 75 clothes washers was conducted with the refrigerator results below. The Energy Star label, size, and price are all positively correlated.

Store	%	price	(ft^3)
Best Buy	53%	2468.99	25.56
Sears	44%	3137.32	25.12
Home Depot	67%	2669.83	25.25
Lowe's	49%	2488.14	25.76
hhgregg	64%	2591.37	25.25
Total	56%	2640.10	25.38

Energy Star refrigerators

Store	%	price	(ft^3)
Best Buy	47%	1029.51	21.71
Sears	56%	2250.79	21.63
Home Depot	33%	1315.33	20.62
Lowe's	51%	1097.96	20.74
hhgregg	36%	1166.87	21.84
Total	44%	1429.72	21.11

Not Energy Star refrigerators

69% of clothes washers on the market are Energy Star labeled. According to RECs 2009 data, 79.72% of households chose a primary Energy Star refrigerator and 76.62% of households chose an Energy Star clothes washer. As these percentages are higher than the availability found in 2015, people shopping for these appliances are actively choosing Energy Star models.

RECs Regression Results

sociodemographic variables	Choose ES Fridge	Have ES Fridge	Choose ES Washer	Have ES Washer	Saves Heat at Night	Efficient Lightbulbs
live with partner	pos***	pos***	pos***	pos***	pos***	pos***
people in house						pos***
sex, female=1	neg**			pos**		
householder education	pos*		pos**		pos***	pos***
urban, yes =1		neg*	neg**			
housholder's age^2		neg***	neg***	neg***		
household income	pos***	pos***	pos**	pos***		
houseage: 1950-1959						
houseage: 1960-1969						
houseage: 1970-1979						neg**
houseage: 1980-1989					neg*	
houseage: 1990-1999						neg***
houseage: 2000-2009		pos***		pos***	neg***	neg***
large apartment complex				neg**		
small apartment complex	neg*	pos*	pos**			
attached house		pos**	pos***	pos**		pos**
separate house		pos***	pos***	pos***		pos***
rents home	neg***	neg***	neg**	neg***	neg***	neg***
free housing				neg***		
Householder is Asian				neg**		
Householder is Black		pos***			pos***	neg***
Householder is Hispanic	neg*	pos*			pos**	neg***
Householder is Pacific Islander						
Householder is Native American						
Householder is another race					neg***	
Householder is mixed race			neg*			
Middle Atlantic						neg***
East North Central		neg***			neg***	neg*
West North Central	neg***	neg***	neg**	neg***	neg**	
South Atlantic	neg***	neg***				
East South Central					neg*	
West South Central			neg**			neg**
Mountain North		neg**				
Mountain South	neg*	neg**				
Pacific		pos***			neg***	pos***

These results are estimated via OLS regressions. Sociodemographic variables that were insignificant in all of the energy efficient factors, such as presence of kids, were excluded from the regressions. Different sociodemographic factors were available in AHS. Both citizenship and a more comprehensive metro variable were included in the AHS regressions.

Other RECs energy efficient factors include choosing an Energy Star dishwasher, having an Energy Star dishwasher, and saving air conditioning during the night. The results of these are less informative and therefore have less value. AHS 2011 energy efficient factors is having an Energy Star refrigerator, clothes washer, and dishwasher. AHS also includes a question about energy efficient jobs performed on houses that spans multiple survey years.

Who is Energy Efficient?

While significance and impact of the sociodemographic factors on household energy efficiency often differs with each variable in both studies, some patterns emerge. Almost unanimous is the importance of the householder living with a partner/spouse and attaining a higher level of education.

Also predictive of energy efficiency is a higher household income and living in a house, regardless of whether that house is attached or detached. Newer homes are more likely to have Energy Star appliances, but house age has no impact on householder's choice to purchase Energy Star appliances. Newer houses have negative significance for other energy efficient measures.

While the householder's squared age is negatively statistically significant, the coefficients in the regression make it economically insignificant. Understandably, renting a home makes energy efficiency less likely.

West North Central and East North Central census divisions are least likely to have energy efficient households. Middle Atlantic and Pacific census divisions are more likely to have energy efficient households.

The impact of race is much more varied than previous studies suggest. Blacks and Hispanics were less likely to install energy efficient light bulbs, but are more likely to have an Energy Star refrigerator and save heat during the night. Race had no significance for any of the energy efficient factors in the AHS. However, AHS demonstrates that those who are not born as US citizens are less likely to be energy efficient.

A resident of a home attending religious services proved to have a positive statistically significant impact on the performance of energy efficient jobs on resident in AHS 2013.

Future Needs

While this analysis is more comprehensive than many before it, having a study targeting energy efficiency would be useful. There are many more factors of energy efficiency that were not covered by these surveys. Also, knowing how aware people are of energy efficient behaviors and motivations behind engaging in them could be very instrumental in creating effective marketing and policies.

There are also specific questions raised by this analysis. Around 20% of AHS respondents refused to answer many of the questions. Depending what sociodemographic groups these households belonged to could greatly impact the AHS results. The differences between RECs and AHS results need to be studied. Finally, the impact of religion, political affiliation, community on energy efficiency should be considered in future analyses.

References

[1] U.S. Energy Information Administration. 2009 RECS Survey Data. website: http://www.eia.gov/consumption/residential/data/2009/
[2] United States Census Bureau. American Housing Survey (AHS). website: http://www.census.gov/programs-surveys/ahs.html
[3] International Energy Agency. 2015. Energy Efficiency. website: http://www.iea.org/topics/energyefficiency/
[4] ES appliance availability: bestbuy.com, sears.com, homedepot.com, lowes.com, & hhgregg.com
[5] images: cdc.gov, illinoisgoc.com, & americanliquidwaste.com

Acknowledgments

- National Science Foundation award DMR-1461275, REU Site: Research Experiences for Undergraduates in Renewable Energy
- Renewable Energy Materials Research Science and Engineering Center REU
- Colorado School of Mines

Year: 2015
Tool: Latex

Worcester Polytechnic Institute

Sociodemographic Factors Influencing Household Energy Efficiency in the United States

ML Tlachac

Motivation

Knowing which people are more energy efficient is increasingly important as the negative effects of local pollution, global warming, and energy insecurity are becoming more pronounced. Determining the sociodemographic factors of the households which are more energy efficient allows for better targeting of marketing and policies.

Energy efficiency is reducing the amount of energy needed to perform a task

Data

This analysis incorporates multiple studies: the Residential Energy Consumption Survey (RECs) 2009 and the American Housing Survey (AHS) 2011 are used for the analysis. RECs 2009 surveyed 12083 households and AHS 2011 surveyed 68162 households within the United States. RECs include more questions about energy efficiency while AHS includes more sociodemographic information. This analysis studies multiple aspects of energy efficiency available in both surveys.

There are two different energy efficient factors for each Energy Star appliance: people can actively choose to purchase an Energy Star appliance and they can have an Energy Star appliance. These have very different implications for marketing and policies.

Energy Star appliance availability was unavailable in the literature. As such, a study incorporating 225 refrigerators, 100 dishwashers, and 75 clothes washers was conducted with the refrigerator results below. The Energy Star label, size, and price are all positively correlated.

Store	%	price	(ft^3)
Best Buy	53%	2468.99	25.56
Sears	44%	3137.32	25.12
Home Depot	67%	2669.83	25.25
Lowe's	49%	2488.14	25.76
hhgregg	64%	2591.37	25.25
Total	56%	2640.10	25.38

Energy Star refrigerators

Store	%	price	(ft^3)
Best Buy	47%	1029.51	21.71
Sears	56%	2250.79	21.63
Home Depot	33%	1315.33	20.62
Lowe's	51%	1097.96	20.74
hhgregg	36%	1166.87	21.84
Total	44%	1429.72	21.11

Not Energy Star refrigerators

69% of clothes washers on the market are Energy Star labeled. According to RECs 2009 data, 79.72% of households chose a primary Energy Star refrigerator and 76.62% of households chose an Energy Star clothes washer. As these percentages are higher than the availability found in 2015, people shopping for these appliances are actively choosing Energy Star models.

RECs Regression Results

sociodemographic variables	Choose ES Fridge	Have ES Fridge	Choose ES Washer	Have ES Washer	Saves Heat at Night	Efficient Lightbulbs
live with partner	pos***	pos***	pos***	pos***	pos***	pos***
people in house						pos***
sex, female=1	neg**			pos**		
householder education	pos*		pos**		pos***	pos***
urban, yes =1		neg*	neg**			
housholder's age^2		neg***	neg***	neg***		
household income	pos***	pos***	pos**	pos***		
houseage: 1950-1959						
houseage: 1960-1969						
houseage: 1970-1979						neg**
houseage: 1980-1989					neg*	
houseage: 1990-1999						neg***
houseage: 2000-2009		pos***		pos***	neg***	neg***
large apartment complex				neg**		
small apartment complex	neg*	pos*	pos**			
attached house		pos**	pos***	pos**		pos**
separate house		pos***	pos***	pos***		pos***
rents home	neg***	neg***	neg**	neg***	neg***	neg***
free housing				neg***		
Householder is Asian				neg**		
Householder is Black		pos***			pos***	neg***
Householder is Hispanic	neg*	pos*			pos**	neg***
Householder is Pacific Islander						
Householder is Native American						
Householder is another race					neg***	
Householder is mixed race			neg*			
Middle Atlantic						neg***
East North Central		neg***			neg***	neg*
West North Central	neg***	neg***	neg**	neg***	neg**	
South Atlantic	neg***	neg***				
East South Central					neg*	
West South Central			neg**			neg**
Mountain North		neg**				
Mountain South	neg*	neg**				
Pacific		pos***			neg***	pos***

These results are estimated via OLS regressions. Sociodemographic variables that were insignificant in all of the energy efficient factors, such as presence of kids, were excluded from the regressions. Different sociodemographic factors were available in AHS. Both citizenship and a more comprehensive metro variable were included in the AHS regressions.

Other RECs energy efficient factors include choosing an Energy Star dishwasher, having an Energy Star dishwasher, and saving air conditioning during the night. The results of these are less informative and therefore have less value. AHS 2011 energy efficient factors is having an Energy Star refrigerator, clothes washer, and dishwasher. AHS also includes a question about energy efficient jobs performed on houses that spans multiple survey years.

Who is Energy Efficient?

While significance and impact of the sociodemographic factors on household energy efficiency often differs with each variable in both studies, some patterns emerge. Almost unanimous is the importance of the householder living with a partner/spouse and attaining a higher level of education.

Also predictive of energy efficiency is a higher household income and living in a house, regardless of whether that house is attached or detached. Newer homes are more likely to have Energy Star appliances, but house age has no impact on householder's choice to purchase Energy Star appliances. Newer houses have negative significance for other energy efficient measures.

While the householder's squared age is negatively statistically significant, the coefficients in the regression make it economically insignificant. Understandably, renting a home makes energy efficiency less likely.

West North Central and East North Central census divisions are least likely to have energy efficient households. Middle Atlantic and Pacific census divisions are more likely to have energy efficient households.

The impact of race is much more varied than previous studies suggest. Blacks and Hispanics were less likely to install energy efficient light bulbs, but are more likely to have an Energy Star refrigerator and save heat during the night. Race had no significance for any of the energy efficient factors in the AHS. However, AHS demonstrates that those who are not born as US citizens are less likely to be energy efficient.

A resident of a home attending religious services proved to have a positive statistically significant impact on the performance of energy efficient jobs on resident in AHS 2013.

Future Needs

While this analysis is more comprehensive than many before it, having a study targeting energy efficiency would be useful. There are many more factors of energy efficiency that were not covered by these surveys. Also, knowing how aware people are of energy efficient behaviors and motivations behind engaging in them could be very instrumental in creating effective marketing and policies.

There are also specific questions raised by this analysis. Around 20% of AHS respondents refused to answer many of the questions. Depending what sociodemographic groups these households belonged to could greatly impact the AHS results. The differences between RECs and AHS results need to be studied. Finally, the impact of religion, political affiliation, community on energy efficiency should be considered in future analyses.

References

[1] U.S. Energy Information Administration. 2009 RECS Survey Data. website: http://www.eia.gov/consumption/residential/data/2009/
[2] United States Census Bureau. American Housing Survey (AHS). website: http://www.census.gov/programs-surveys/ahs.html
[3] International Energy Agency. 2015. Energy Efficiency. website: http://www.iea.org/topics/energyefficiency/
[4] ES appliance availability: bestbuy.com, sears.com, homedepot.com, lowes.com, & hhgreg.com
[5] images: cdc.gov, illinoisgoc.com, & americanliquidwaste.com

Acknowledgments

- National Science Foundation award DMR-1461275, REU Site: Research Experiences for Undergraduates in Renewable Energy
- Renewable Energy Materials Research Science and Engineering Center REU
- Colorado School of Mines

Year: 2015
Tool: Latex

Self Critiques

- Colored text in table difficult to read
- Too much text
- Text too small
- Can't match references
- Good use of tables to break up walls of text

Worcester Polytechnic Institute

Year: 2015
Tool: Latex

Statistical Analysis of Crime in Eau Claire, WI

Matthew Tlachac[†], ML Tlachac[*]

Research Advisor: Abra Brisbin[*]

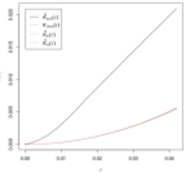
[*]University of Wisconsin-Eau Claire, [†]University of Minnesota

Motivation and Data

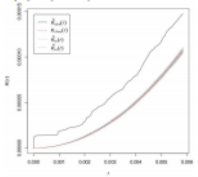
When moving off campus or to a new city, safety from crime is often a primary concern. Using crime data, we use statistics to identify factors that influence crime distribution in Eau Claire, WI. This could potentially help people determine where they would like to reside. Crimes occurring during the last month of spring semester 2015 and the first month of fall semester 2015 were collected from crimemapping.com[1].

Overall, during the 8 weeks there was a total of 1124 crimes. The scatter plot and results of Ripley's K-function for these crimes are below. The Ripley's K-function, using the isomorphic translation, gives evidence that the crime distribution is not random over a homogeneous Poisson distribution. Since a random distribution of crimes is very unlikely given the geography of Eau Claire, we have focused on a restricted area which we deem "downtown". The Ripley's K-function still shows the crimes are clustered. The downtown area will be used for the remainder of the analyses.

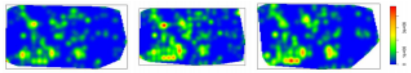
Ripley's K-function for crimes in Eau Claire



Ripley's K-function for crimes in downtown

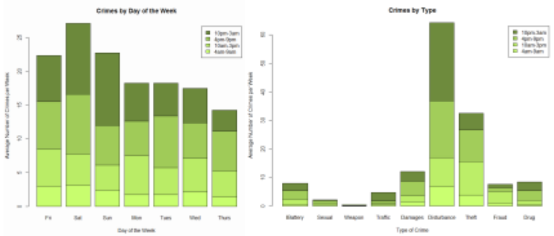


575 crimes occurred during the end of the spring semester and 549 during the beginning of the fall semester. Weather has been shown to impact crime, but May and September have very similar average high and low temperatures. The below density plots demonstrate the difference in locations of crime in the downtown area based on semester. Whether the difference is due to a seasonal or yearly effect is unknown given the limited data. Ripley's K-function yielded that crimes in the spring and fall are clustered.

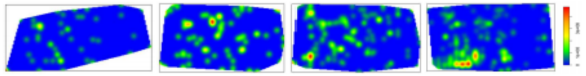


Fall downtown crimes *All downtown crimes* *Spring downtown crimes*

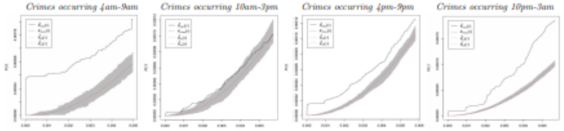
Temporal Analysis



We choose to focus on four different time periods to determine if time was influential on crime. The time periods are 4am-9am (morning), 10am-3pm (day), 4pm-9pm (evening), and 10pm-3am (night). We used Pearson's Chi-squared tests to analyze the information in the above bar plots. There is an association between the time a crime occurs and the day of the week (p-value $= 2.45 \cdot 10^{-4}$). In particular, crimes on the weekend occur at different times than crimes on weekdays (p-value $= 7.85 \cdot 10^{-5}$). Also, there is an association between the type of crime and when it occurs (p-value $= 7.49 \cdot 10^{-16}$). Thus, an individual's daily and weekly schedule may impact where they want to reside. Given that, we have looked at the spatial crime distribution during each time period using density maps. From Ripley's K-function crimes are not distributed randomly during 4am-9am, 4pm-9pm, and 10pm-3am.

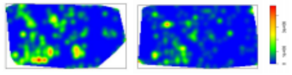


Crimes occurring 4am-9am *Crimes occurring 10am-3pm* *Crimes occurring 4pm-9pm* *Crimes occurring 10pm-3am*



Broken Window Hypothesis

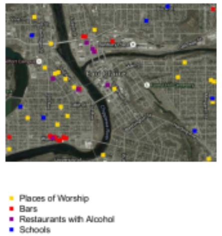
The broken window hypothesis is a theory stating that "crime emanated from disorder and that if disorder were eliminated, then serious crimes would not occur"[3]. As such, we have isolated the disturbances to compare against the other crimes. The difference in crime distribution in the resulting density plots could be interpreted in multiple ways. First, since we would expect the distribution of disturbances to predict the distribution of other crimes, we could say the hypothesis is false. Alternatively, since the disturbances were reported, we would expect there to be fewer serious crimes in that locale, indicating the hypothesis is true.

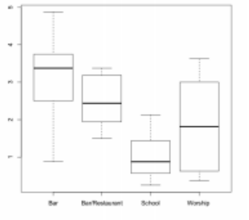


Disturbance crimes *All other crimes*

Locations of Interest

A study conducted by the University of New Mexico found that there is a relation between types of establishments and crimes[2]. They found violent, property, and drug crime were more common around blocks with bars, restaurants, and high schools; churches and elementary schools had no notable relation. We looked at all crimes within a radius of 0.0025 degrees of latitude and longitude around the 16 bars, 8 restaurants with alcohol, 7 schools, and 22 places of worship in downtown Eau Claire.





Number of crimes per week at given locations

It was unexpected to find that schools had a lower mean number of crimes than locations of worship. The higher mean number of crimes around bars was expected but the impact of multiple bars on Water Street needs to be further analyzed.

Future Research

This research project is currently in progress. There is still much analysis left to complete. Our analyses would be strengthened by collecting additional data. We have identified important features of the relationship between time, location, and type of crime. In the future, we plan to extend this analysis in the following ways.

- Determine if the difference of seasonal crime patterns persists with more seasons. Additionally, see if there is a difference in yearly crime patterns.
- Further explore the different interpretations of the broken window hypothesis.
- Investigate the frequency of crimes around bars, schools, and places of worship in all of Eau Claire. Other establishments than those in the New Mexico study could also be included in analysis.
- Analyze how the number of crimes near stated establishments compare to the number of crimes near random locations in Eau Claire.
- Repeat all of the analyses completed with each individual type of crime. That way depending on the types of crimes that cause concern, people could choose where moved more selectively.

This research could be extended to any city. The results would be more impactful in a city with more violent crimes. This type of research could help advise people when the move off campus or to a new city so they can find the most desirable place to live.

References

[1] Crimemapping.com. (2015). The Omega Group Inc.

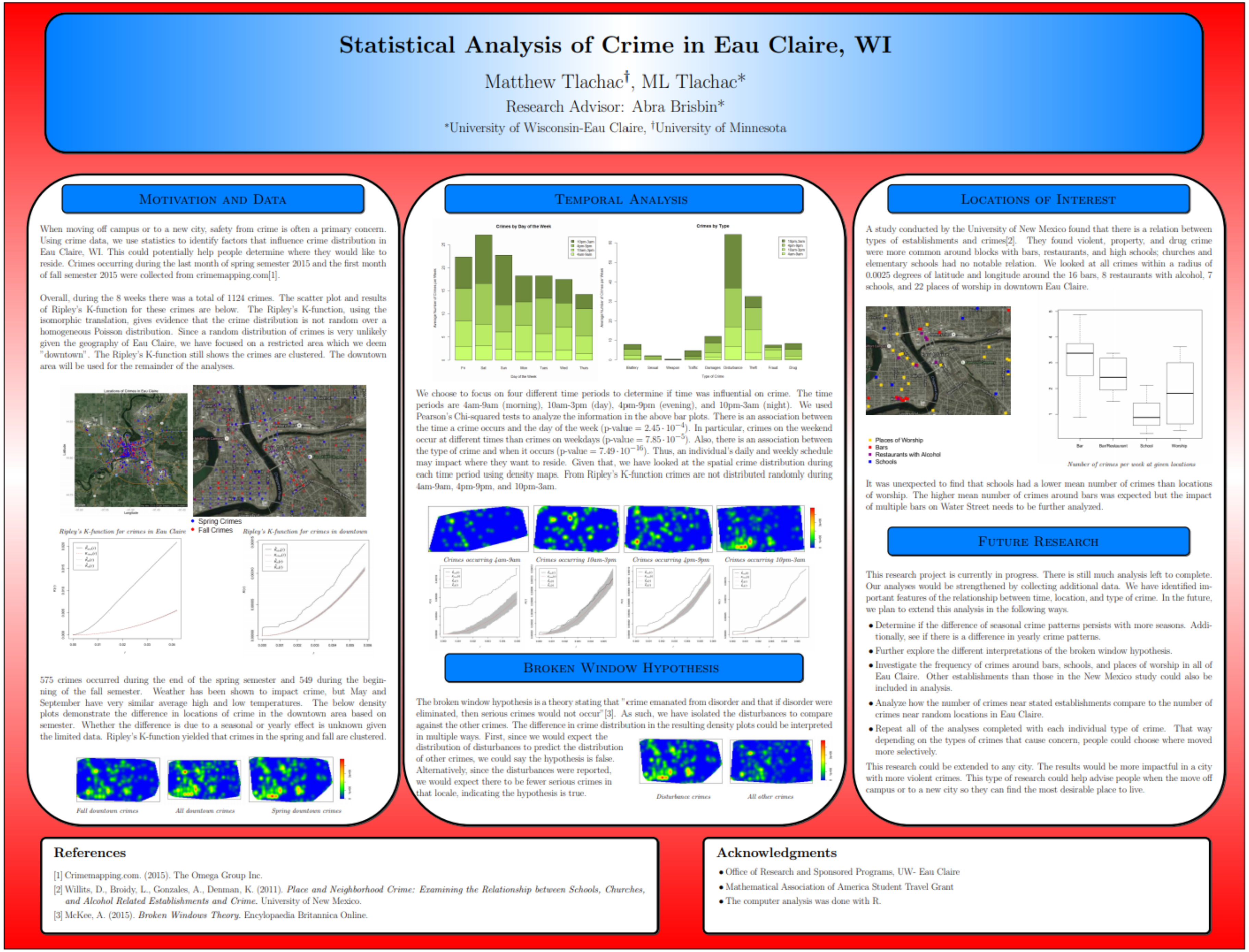
[2] Willits, D., Broidy, L., Gonzales, A., Denman, K. (2011). *Place and Neighborhood Crime: Examining the Relationship between Schools, Churches, and Alcohol Related Establishments and Crime.* University of New Mexico.

[3] McKee, A. (2015). *Broken Windows Theory.* Encyclopaedia Britannica Online.

Acknowledgments

- Office of Research and Sponsored Programs, UW- Eau Claire
- Mathematical Association of America Student Travel Grant
- The computer analysis was done with R.

Worcester Polytechnic Institute



Statistical Analysis of Crime in Eau Claire, WI

Matthew Tlachac[†], ML Tlachac[*]
Research Advisor: Abra Brisbin[*]
[*]University of Wisconsin-Eau Claire, [†]University of Minnesota

Motivation and Data

When moving off campus or to a new city, safety from crime is often a primary concern. Using crime data, we use statistics to identify factors that influence crime distribution in Eau Claire, WI. This could potentially help people determine where they would like to reside. Crimes occurring during the last month of spring semester 2015 and the first month of fall semester 2015 were collected from crimemapping.com[1].

Overall, during the 8 weeks there was a total of 1124 crimes. The scatter plot and results of Ripley's K-function for these crimes are below. The Ripley's K-function, using the isomorphic translation, gives evidence that the crime distribution is not random over a homogeneous Poisson distribution. Since a random distribution of crimes is very unlikely given the geography of Eau Claire, we have focused on a restricted area which we deem "downtown". The Ripley's K-function still shows the crimes are clustered. The downtown area will be used for the remainder of the analyses.

Spring Crimes
Fall Crimes

Ripley's K-function for crimes in Eau Claire *Ripley's K-function for crimes in downtown*

575 crimes occurred during the end of the spring semester and 549 during the beginning of the fall semester. Weather has been shown to impact crime, but May and September have very similar average high and low temperatures. The below density plots demonstrate the difference in locations of crime in the downtown area based on semester. Whether the difference is due to a seasonal or yearly effect is unknown given the limited data. Ripley's K-function yielded that crimes in the spring and fall are clustered.

Fall downtown crimes *All downtown crimes* *Spring downtown crimes*

Temporal Analysis

We choose to focus on four different time periods to determine if time was influential on crime. The time periods are 4am-9am (morning), 10am-3pm (day), 4pm-9pm (evening), and 10pm-3am (night). We used Pearson's Chi-squared tests to analyze the information in the above bar plots. There is an association between the time a crime occurs and the day of the week (p-value $= 2.45 \cdot 10^{-4}$). In particular, crimes on the weekend occur at different times than crimes on weekdays (p-value $= 7.85 \cdot 10^{-5}$). Also, there is an association between the type of crime and when it occurs (p-value $= 7.49 \cdot 10^{-16}$). Thus, an individual's daily and weekly schedule may impact where they want to reside. Given that, we have looked at the spatial crime distribution during each time period using density maps. From Ripley's K-function crimes are not distributed randomly during 4am-9am, 4pm-9pm, and 10pm-3am.

Crimes occurring 4am-9am *Crimes occurring 10am-3pm* *Crimes occurring 4pm-9pm* *Crimes occurring 10pm-3am*

Broken Window Hypothesis

The broken window hypothesis is a theory stating that "crime emanated from disorder and that if disorder were eliminated, then serious crimes would not occur"[3]. As such, we have isolated the disturbances to compare against the other crimes. The difference in crime distribution in the resulting density plots could be interpreted in multiple ways. First, since we would expect the distribution of disturbances to predict the distribution of other crimes, we could say the hypothesis is false. Alternatively, since the disturbances were reported, we would expect there to be fewer serious crimes in that locale, indicating the hypothesis is true.

Disturbance crimes *All other crimes*

Locations of Interest

A study conducted by the University of New Mexico found that there is a relation between types of establishments and crimes[2]. They found violent, property, and drug crime were more common around blocks with bars, restaurants, and high schools; churches and elementary schools had no notable relation. We looked at all crimes within a radius of 0.0025 degrees of latitude and longitude around the 16 bars, 8 restaurants with alcohol, 7 schools, and 22 places of worship in downtown Eau Claire.

Places of Worship
Bars
Restaurants with Alcohol
Schools

Number of crimes per week at given locations

It was unexpected to find that schools had a lower mean number of crimes than locations of worship. The higher mean number of crimes around bars was expected but the impact of multiple bars on Water Street needs to be further analyzed.

Future Research

This research project is currently in progress. There is still much analysis left to complete. Our analyses would be strengthened by collecting additional data. We have identified important features of the relationship between time, location, and type of crime. In the future, we plan to extend this analysis in the following ways.

- Determine if the difference of seasonal crime patterns persists with more seasons. Additionally, see if there is a difference in yearly crime patterns.
- Further explore the different interpretations of the broken window hypothesis.
- Investigate the frequency of crimes around bars, schools, and places of worship in all of Eau Claire. Other establishments than those in the New Mexico study could also be included in analysis.
- Analyze how the number of crimes near stated establishments compare to the number of crimes near random locations in Eau Claire.
- Repeat all of the analyses completed with each individual type of crime. That way depending on the types of crimes that cause concern, people could choose where moved more selectively.

This research could be extended to any city. The results would be more impactful in a city with more violent crimes. This type of research could help advise people when the move off campus or to a new city so they can find the most desirable place to live.

References

[1] Crimemapping.com. (2015). The Omega Group Inc.
[2] Willits, D., Broidy, L., Gonzales, A., Denman, K. (2011). *Place and Neighborhood Crime: Examining the Relationship between Schools, Churches, and Alcohol Related Establishments and Crime.* University of New Mexico.
[3] McKee, A. (2015). *Broken Windows Theory.* Encyclopaedia Britannica Online.

Acknowledgments

- Office of Research and Sponsored Programs, UW- Eau Claire
- Mathematical Association of America Student Travel Grant
- The computer analysis was done with R.

Year: 2015
Tool: Latex

Self Critiques

- Bad color choice
- Missing header logos
- Too much text
- Text too small
- Too much future research
- Good inclusion of visualizations

Stage 2: Early Grad

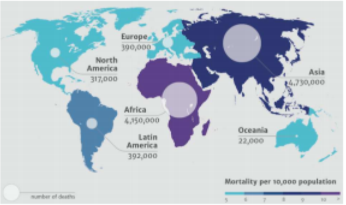
Year: 2018
Tool: Latex

Tackling the Antibiotic Resistance Crisis using Longitudinal Massachusetts Antibiogram Data

Student: ML Tlachac*, Advisor: Elke Rundensteiner*
Medical Collaborators: Kerri Barton†, Scott Troppy†, Kirthana Beaulac‡, Shira Doron‡
*Worcester Polytechnic Institute (WPI), †Massachusetts Department of Public Health (MDPH), ‡Tufts Medical Center

Motivation

Antibiotic resistant bacteria become more prevalent every year. In 2013 in the U.S., they caused **23 thousand deaths** and **$40 billion** [1].

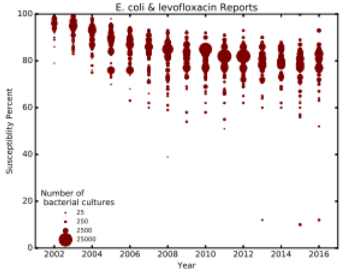


2050 forecasts of antibiotic resistance deaths [2]

Responsibly prescribing antibiotics can mitigate this growing resistance. This requires predictions of the current antibiotic susceptibility.

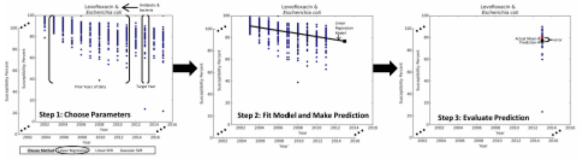
Hospital Antibiogram Data

The MDPH collected statewide susceptibility reports for 15 years [3,4].

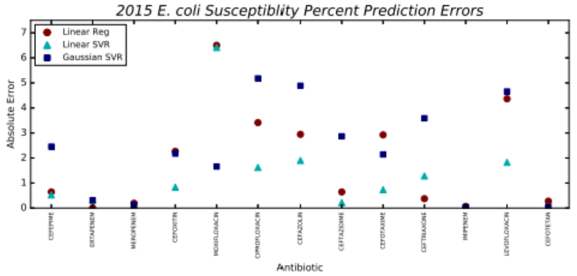


Over **16 thousand** reports containing *E. coli*, *Klebsiella oxytoca*, and *Klebsiella pneumoniae* susceptibility to carbapenems, cephalosporins, and fluoroquinolones [5].

Prediction Methodology

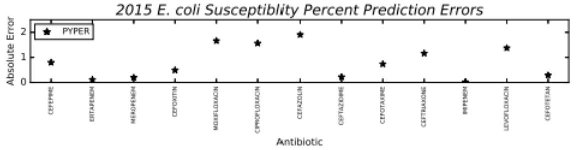


We predict the mean susceptibility percent in 2014, 2015, and 2016 using the prior 12 years of data. Methods include linear regression as well as linear and Gaussian support vector regression (SVR).



PYPER Model Selector

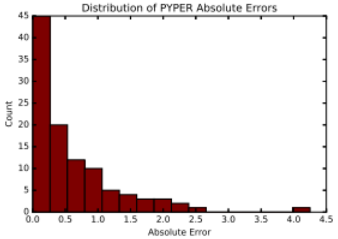
Different methods are better for different antibiotic-bacteria pairs. PYPER chooses a method to use based on which model predicted the susceptibility of the previous year best [3].



Prediction Evaluation

Mean absolute error $= \frac{1}{m}\sum_{i=1}^{m}|p_i - a_i|$ for m antibiotic-bacteria pairs where p_i is the predicted and a_i is the actual mean susceptibility percent.

Method	Mean Absolute Error
Linear Regression	1.32
Linear SVR	1.13
Gaussian SVR	1.20
PYPER	**0.62**



Conclusion and Impact

On average, PYPER predicted the future susceptibility percent within 0.62% of the actual susceptibility percent. These predictions can

- guide prescription practices
- create treatment policies
- guide research and development
- prepare for antibiotic resistant infections

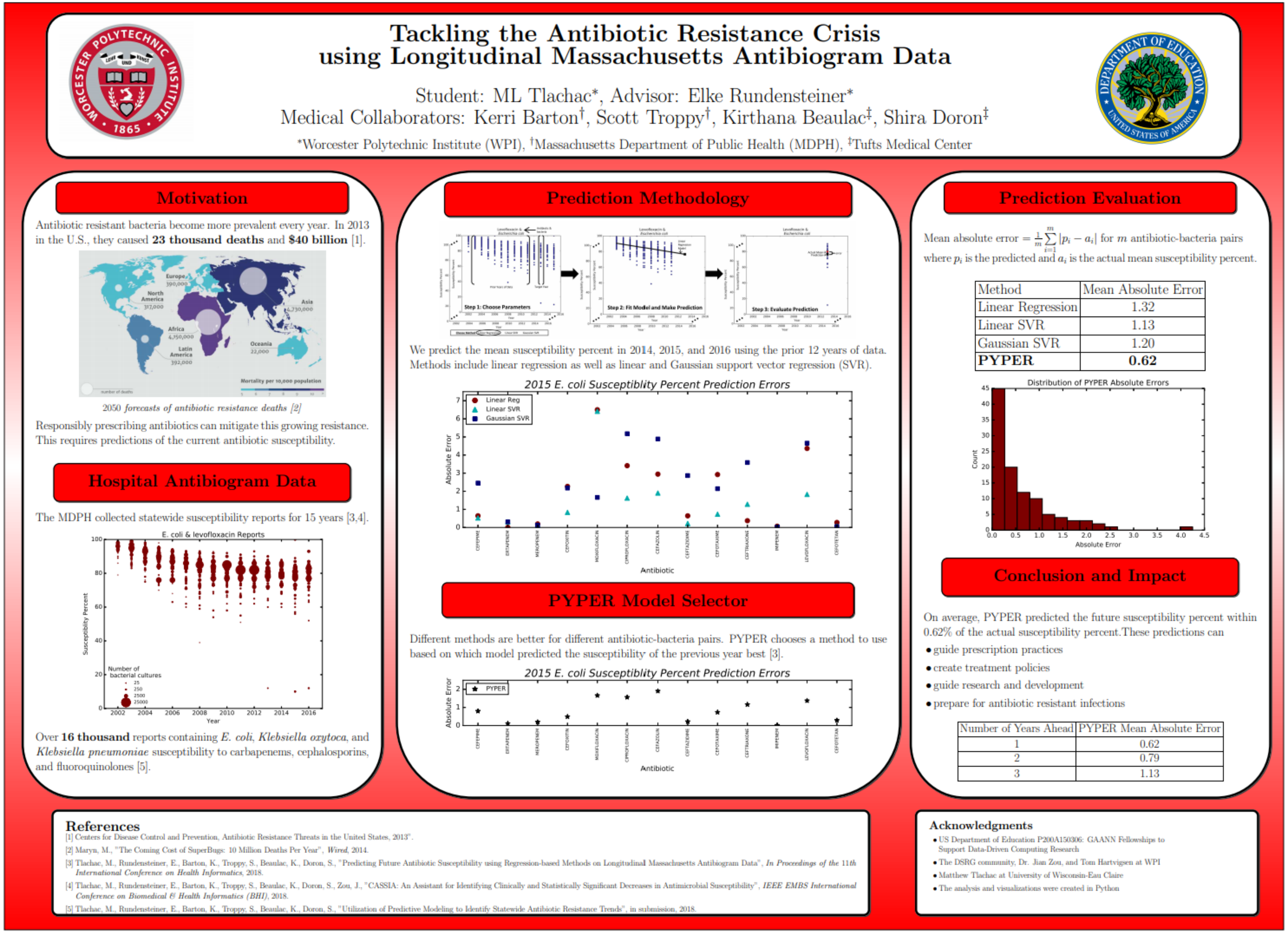
Number of Years Ahead	PYPER Mean Absolute Error
1	0.62
2	0.79
3	1.13

References

[1] Centers for Disease Control and Prevention, Antibiotic Resistance Threats in the United States, 2013".
[2] Maryn, M., "The Coming Cost of SuperBugs: 10 Million Deaths Per Year", *Wired*, 2014.
[3] Tlachac, M., Rundensteiner, E., Barton, K., Troppy, S., Beaulac, K., Doron, S., "Predicting Future Antibiotic Susceptibility using Regression-based Methods on Longitudinal Massachusetts Antibiogram Data", *In Proceedings of the 11th International Conference on Health Informatics*, 2018.
[4] Tlachac, M., Rundensteiner, E., Barton, K., Troppy, S., Beaulac, K., Doron, S., Zou, J., "CASSIA: An Assistant for Identifying Clinically and Statistically Significant Decreases in Antimicrobial Susceptibility", *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2018.
[5] Tlachac, M., Rundensteiner, E., Barton, K., Troppy, S., Beaulac, K., Doron, S., "Utilization of Predictive Modeling to Identify Statewide Antibiotic Resistance Trends", in submission, 2018.

Acknowledgments

- US Department of Education P200A150306: GAANN Fellowships to Support Data-Driven Computing Research
- The DSRG community, Dr. Jian Zou, and Tom Hartvigsen at WPI
- Matthew Tlachac at University of Wisconsin-Eau Claire
- The analysis and visualizations were created in Python



Year: 2018
Tool: Latex

Self Critiques

- First figure in middle panel is too small
- Good impacts

Year: 2019
Tool: Latex

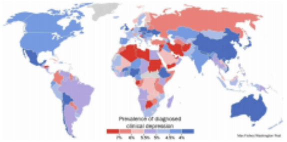
Predicting Mental Health from Smartphone Text Messages

Student: ML Tlachac, Advisor: Elke Rundensteiner

Worcester Polytechnic Institute

Motivation

Depression, a prevalent and debilitating mental illness, is frequently undiagnosed. Despite being one of the most treatable mental disorders, it is the leading cause of disability in adults [1,2].



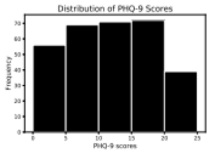
Diagnosed depression rates in 2013 [3].

Opportunity: screen for depression with Smartphone data.

Dataset & Research Questions

A team at WPI collected two weeks of retrospectively harvested Smartphone data [4]. There are text messages and Patient Health Questionnaire-9 (PHQ-9) scores for 314 participants. The PHQ-9 is a tool used to screen, diagnose, and monitor depression [1].

PHQ-9 Score	Interim Diagnosis	Recommendation
$0-4$	Not Depressed	
$5-9$	Minimally Symptomatic	Watch
$10-14$	Mild Depression	Treatment
$15-19$	Moderate Depression	Treatment
20+	Severe Depression	Treatment



Q1: Are received texts a useable modality to predict PHQ-9 scores?
Q2: Are features generated from a subset of contacts better than features from all contacts?

Contact Subsets

The number of total contacts n for each participant P are between 1 and 420 contacts. The contact subsets for each participant P are:

- Top 1 contact
- Top $C_P(a) = a$ contacts for $a \in \{2, 3, 4\}$
- Top $C_P(r) = r(n)$ contacts for $r \in \{25\%, 50\%, 75\%\}$
- Top $C_P(w) = \sum_{i=1}^{n}(F_i)$ for $F_i = \begin{cases} 1 \; if \; t_i \geq w(max(T)) \\ 0 \; otherwise \end{cases}$ contacts for $w \in \{0.75, 0.5, 0.25\}$

where t_i is the ith contact's number of texts and T is the set of the number of texts for P.

- All contacts

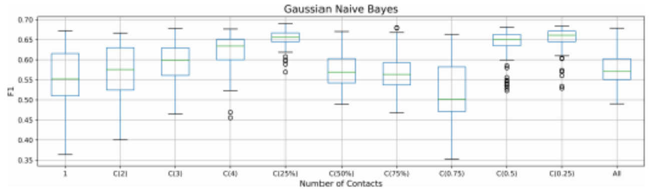
Text Feature Engineering

The quantity and quality of social interactions are known predictors of health [6]. For each subset of contacts for each participant, we generate:

- Polarity features: percent positive, percent negative, average positivity, average negativity
- Subjective features: percent subjective, average subjectivity
- Part of speech (POS) features: percent of POS tag for all 36 tags
- Volume features: number of contacts, number of texts, average number of POS tags/text

Depression Prediction

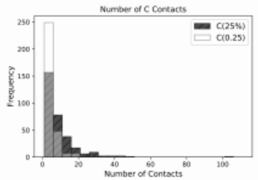
We predict depression using a binary PHQ-9 score of cutoff 10 with 5-fold cross-validation.



Features from the top $C_P(25\%)$ or $C_P(0.25)$ contacts improved the F1 score by 13.2 percent to 0.653. This F1 score increase of 0.577 is statistically significant.

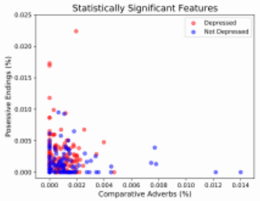
Number of Contacts

The top $C_P(25\%)$ and $C_P(0.25)$ contacts yielded the best results.



Feature Importance

Feature	p-value	Impact
Comparative Adverbs (%)	$p = 0.001$	Decreases Depression Score
Possessive Endings (%)	$p = 0.018$	Increases Depression Score



Conclusion & Impact

A1. Received communications could be a useful modality, particularity when combined with other modalities.

A2. Focusing on a subset of contacts is a promising approach that could be deployed in any study leveraging received communications.

In submission to IEEE International Conference on Biomedical and Health Informatics

References

[1] Kroenke, et al. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9).
[2] Siu, et al. (2016). Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*, 315(4).
[3] Dewey. (2013). *A stunning map of depression rates around the world*. Washington Post.
[4] Dogrucu, et al. (2019). Instantaneous Depression Assessment using Machine Learning on Voice Samples and Retrospectively Harvested Smart-phone and Social Media Data. In Submission.
[5] Center for Quality Assessment and Improvement in Mental Health. (2017). Patient Health Questionairre (PHQ-9).
[6] Cohen. (2004). Social relationships and health. *American Psychological Association*, 59(8).

Acknowledgments

- US Department of Education P200A150306: GAANN Fellowships to Support Data-Driven Computing Research
- Ermal Toto
- The DSRG community
- Prof. Agu, Resom, Flannery, Dogrucu, Peruic, Isaro, Damon, Gao, Wu, Assan

Worcester Polytechnic Institute

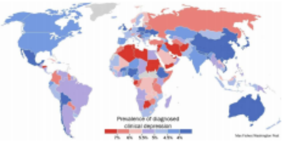
Predicting Mental Health from Smartphone Text Messages

Student: ML Tlachac, Advisor: Elke Rundensteiner

Worcester Polytechnic Institute

Motivation

Depression, a prevalent and debilitating mental illness, is frequently undiagnosed. Despite being one of the most treatable mental disorders, it is the leading cause of disability in adults [1,2].



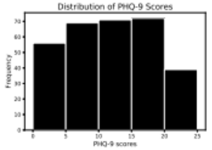
Diagnosed depression rates in 2013 [3].

Opportunity: screen for depression with Smartphone data.

Dataset & Research Questions

A team at WPI collected two weeks of retrospectively harvested Smartphone data [4]. There are text messages and Patient Health Questionnaire-9 (PHQ-9) scores for 314 participants. The PHQ-9 is a tool used to screen, diagnose, and monitor depression [1].

PHQ-9 Score	Interim Diagnosis	Recommendation
$0-4$	Not Depressed	
$5-9$	Minimally Symptomatic	Watch
$10-14$	Mild Depression	Treatment
$15-19$	Moderate Depression	Treatment
20+	Severe Depression	Treatment



Q1: Are received texts a useable modality to predict PHQ-9 scores?
Q2: Are features generated from a subset of contacts better than features from all contacts?

Contact Subsets

The number of total contacts n for each participant P are between 1 and 420 contacts. The contact subsets for each participant P are:

- Top 1 contact
- Top $C_P(a) = a$ contacts for $a \in \{2,3,4\}$
- Top $C_P(r) = r(n)$ contacts for $r \in \{25\%, 50\%, 75\%\}$
- Top $C_P(w) = \sum_{i=1}^{n}(F_i)$ for $F_i = \begin{matrix} 1\ if\ t_i \geq w(max(T)) \\ 0\ otherwise \end{matrix}$ contacts for $w \in \{0.75, 0.5, 0.25\}$

where t_i is the ith contact's number of texts and T is the set of the number of texts for P.

- All contacts

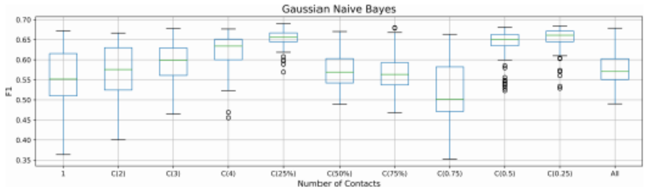
Text Feature Engineering

The quantity and quality of social interactions are known predictors of health [6]. For each subset of contacts for each participant, we generate:

- Polarity features: percent positive, percent negative, average positivity, average negativity
- Subjective features: percent subjective, average subjectivity
- Part of speech (POS) features: percent of POS tag for all 36 tags
- Volume features: number of contacts, number of texts, average number of POS tags/text

Depression Prediction

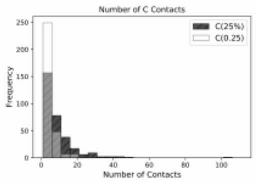
We predict depression using a binary PHQ-9 score of cutoff 10 with 5-fold cross-validation.



Features from the top $C_P(25\%)$ or $C_P(0.25)$ contacts improved the F1 score by 13.2 percent to 0.653. This F1 score increase of 0.577 is statistically significant.

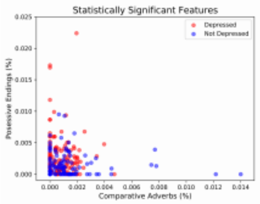
Number of Contacts

The top $C_P(25\%)$ and $C_P(0.25)$ contacts yielded the best results.



Feature Importance

Feature	p-value	Impact
Comparative Adverbs (%)	$p = 0.001$	Decreases Depression Score
Possessive Endings (%)	$p = 0.018$	Increases Depression Score



Conclusion & Impact

A1. Received communications could be a useful modality, particularity when combined with other modalities.

A2. Focusing on a subset of contacts is a promising approach that could be deployed in any study leveraging received communications.

In submission to IEEE International Conference on Biomedical and Health Informatics

References

[1] Kroenke, et al. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9).
[2] Siu, et al. (2016). Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*, 315(4).
[3] Dewey. (2013). *A stunning map of depression rates around the world*. Washington Post.
[4] Dogrucu, et al. (2019). Instantaneous Depression Assessment using Machine Learning on Voice Samples and Retrospectively Harvested Smart-phone and Social Media Data. In Submission.
[5] Center for Quality Assessment and Improvement in Mental Health. (2017). Patient Heatlh Questionairre (PHQ-9).
[6] Cohen. (2004). Social relationships and health. *American Psychological Association*, 59(8).

Acknowledgments

- US Department of Education P200A150306: GAANN Fellowships to Support Data-Driven Computing Research
- Ermal Toto
- The DSRG community
- Prof. Agu, Resom, Flannery, Dogrucu, Peruic, Isaro, Damon, Gao, Wu, Assan

Year: 2019
Tool: Latex

Self Critiques

- Not clear Q1 and Q2 correspond to A1 and A2
- Not all visualizations work well for a poster
- Good use of bullet points

Year: 2019
Tool: Latex

The 10 Most Important Features in Predicting Depression from Content of Retrospectively Harvested Text Messages

Student: ML Tlachac, PhD Candidate & Advisor: Elke Rundensteiner, PhD

Data Science
Worcester Polytechnic Institute

Motivation

Depression, a prevalent and debilitating mental illness, is frequently undiagnosed. Despite being one of the most treatable mental disorders, it is the leading cause of disability in adults [1,2].

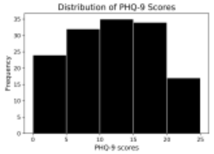
Opportunity: screen for depression with Smartphone data.

- Smartphone data is unobtrusive to collect
- This strategy doesn't require individuals to seek help
- Retrospectively harvested data is biased

Moodable Dataset

A team at WPI collected two weeks of retrospectively harvested Smartphone data, resulting in the Moodable dataset [3]. There are two weeks of sent text messages and Patient Health Questionnaire-9 (PHQ-9) scores for 146 participants. The PHQ-9 is a tool used to screen, diagnose, and monitor depression [1].

PHQ-9 Score	Interim Diagnosis	Recommendation
0 – 4	Not Depressed	
5 – 9	Minimally Symptomatic	Watch
10 – 14	Mild Depression	Treatment
15 – 19	Moderate Depression	Treatment
20+	Severe Depression	Treatment



Our Contributions

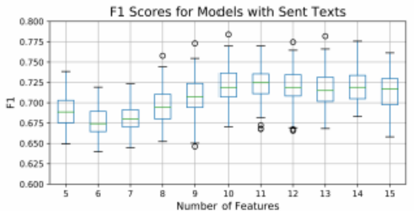
1) Exploring the potential of retrospectively harvested text messages in predicting depression, and
2) Identifying the features from sent text messages that are most influential when predicting depression.

Depression Prediction

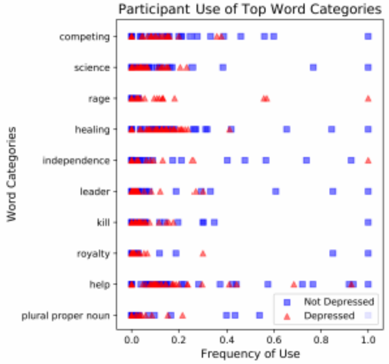
For the text messages from each participant, we generate 210 features.

- Part of speech (POS) features: percent of POS tag for all 36 present tags.
- Word category features: frequency of use of words in 174 lexical categories [4].

We predict depression using a binary PHQ-9 score of cutoff 10 with 5-fold cross-validation. Gaussian Naive Bayes classifiers achieved the highest F1 scores. The top features were selected as those with the highest chi-squared statistic in comparison to the PHQ-9 score.

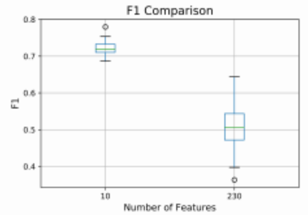


The average F1 score for models with 10 features is 0.721, which is statistically significantly higher than the F1 score of models with fewer features.



Importance of Feature Selection

The average F1 score of models built with the top 10 features is 0.216 higher than the models built with all features.



Case Study: *Rage*

A higher frequency of *rage* words is indicative of depression. While there are 107 words in the *rage* category, the three participants with the highest *rage* word frequencies only use the words "angry" and "furious":

- "Its funny when shes **angry**"
- "**Angry** with you as well"
- "But i should be **angry** because she tricked me [angry emoticon]"
- "She is **angry** today"
- "dnt make me **angry**
- "…He **angry** and depressed because …I would be **furious**"

Conclusion & Impact

This is a preliminary study in analyzing the depression prediction ability of retrospectively harvested sent text messages. Using a computationally inexpensive method, we were able to achieve an F1 score of 0.721 and identify the top 10 most influential features. This research methodology can be deployed in any study containing communications, such as received text messages [5] and tweets [6]. Our future work involves integrating these text features with other Smartphone modalities.

References

[1] Kroenke, et al. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9).
[2] Siu, et al. (2016). Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*, 315(4).
[3] Dogrucu, et al. (2019). Instantaneous Depression Assessment using Machine Learning on Voice Samples and Retrospectively Harvested Smart-phone and Social Media Data. In Submission.
[4] Fast, et al. (2016). Empath: Understanding Topic Signals in Large-Scale Text. In Proceedings of the 2016 *CHI Conference on Human Factors in Computing Systems*.
[5] Tlachac, et al. (2019). You're Making Me Depressed: Leveraging Texts from Contact Subsets to Predict Depression. *IEEE BHI*.
[6] Tlachac, et al. (2019). Tweets as a Diagnostic Tool: Leveraging Twitter Post Content to Predict Depression. In Submission to *IEEE/ACM ASONAM*.

Acknowledgments

- US Department of Education P200A150306: GAANN Fellowships to Support Data-Driven Computing Research
- Ermal Toto and the DSRG community for continued support
- Prof. Agu, Dogrucu, Peruic, Isaro, Ball for innovating the Moodable approach and the resulting dataset
- Resom, Flannery, Gao, Wu, Assan for preliminary analysis

Worcester Polytechnic Institute

The 10 Most Important Features in Predicting Depression from Content of Retrospectively Harvested Text Messages

Student: ML Tlachac, PhD Candidate & Advisor: Elke Rundensteiner, PhD

Data Science
Worcester Polytechnic Institute

Motivation

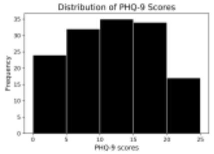
Depression, a prevalent and debilitating mental illness, is frequently undiagnosed. Despite being one of the most treatable mental disorders, it is the leading cause of disability in adults [1,2].

Opportunity: screen for depression with Smartphone data.
- Smartphone data is unobtrusive to collect
- This strategy doesn't require individuals to seek help
- Retrospectively harvested data is biased

Moodable Dataset

A team at WPI collected two weeks of retrospectively harvested Smartphone data, resulting in the Moodable dataset [3]. There are two weeks of sent text messages and Patient Health Questionnaire-9 (PHQ-9) scores for 146 participants. The PHQ-9 is a tool used to screen, diagnose, and monitor depression [1].

PHQ-9 Score	Interim Diagnosis	Recommendation
0 – 4	Not Depressed	
5 – 9	Minimally Symptomatic	Watch
10 – 14	Mild Depression	Treatment
15 – 19	Moderate Depression	Treatment
20+	Severe Depression	Treatment



Our Contributions

1) Exploring the potential of retrospectively harvested text messages in predicting depression, and
2) Identifying the features from sent text messages that are most influential when predicting depression.

Depression Prediction

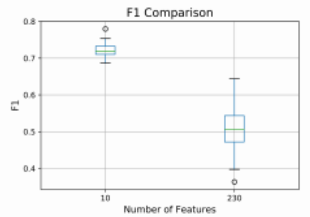
For the text messages from each participant, we generate 210 features.
- Part of speech (POS) features: percent of POS tag for all 36 present tags.
- Word category features: frequency of use of words in 174 lexical categories [4].

We predict depression using a binary PHQ-9 score of cutoff 10 with 5-fold cross-validation. Gaussian Naive Bayes classifiers achieved the highest F1 scores. The top features were selected as those with the highest chi-squared statistic in comparison to the PHQ-9 score.

The average F1 score for models with 10 features is 0.721, which is statistically significantly higher than the F1 score of models with fewer features.

Importance of Feature Selection

The average F1 score of models built with the top 10 features is 0.216 higher than the models built with all features.



Case Study: *Rage*

A higher frequency of *rage* words is indicative of depression. While there are 107 words in the *rage* category, the three participants with the highest *rage* word frequencies only use the words "angry" and "furious":
- "Its funny when shes **angry**"
- "**Angry** with you as well"
- "But i should be **angry** because she tricked me [angry emoticon]"
- "She is **angry** today"
- "dnt make me **angry**"
- "…He **angry** and depressed because …I would be **furious**"

Conclusion & Impact

This is a preliminary study in analyzing the depression prediction ability of retrospectively harvested sent text messages. Using a computationally inexpensive method, we were able to achieve an F1 score of 0.721 and identify the top 10 most influential features. This research methodology can be deployed in any study containing communications, such as received text messages [5] and tweets [6]. Our future work involves integrating these text features with other Smartphone modalities.

References

[1] Kroenke, et al. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9).
[2] Siu, et al. (2016). Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*, 315(4).
[3] Dogrucu, et al. (2019). Instantaneous Depression Assessment using Machine Learning on Voice Samples and Retrospectively Harvested Smart-phone and Social Media Data. In Submission.
[4] Fast, et al. (2016). Empath: Understanding Topic Signals in Large-Scale Text. In Proceedings of the 2016 *CHI Conference on Human Factors in Computing Systems*.
[5] Tlachac, et al. (2019). You're Making Me Depressed: Leveraging Texts from Contact Subsets to Predict Depression. *IEEE BHI*.
[6] Tlachac, et al. (2019). Tweets as a Diagnostic Tool: Leveraging Twitter Post Content to Predict Depression. In Submission to *IEEE/ACM ASONAM*.

Acknowledgments

- US Department of Education P200A150306: GAANN Fellowships to Support Data-Driven Computing Research
- Ermal Toto and the DSRG community for continued support
- Prof. Agu, Dogrucu, Peruic, Isaro, Ball for innovating the Moodable approach and the resulting dataset
- Resom, Flannery, Gao, Wu, Assan for preliminary analysis

Year: 2019
Tool: Latex

Self Critiques

- Empty space
- Could highlight most important content, like in case study
- Good examples with bolded words

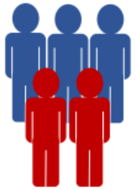
Stage 3: Recent

Year: 2020
Tool: PowerPoint

Depression Screening with Text Messages

ML Tlachac, Data Science PhD Candidate Advisor: Elke Rundensteiner

Motivation



2 in 5 graduate students suffer from **depression**[1].

Despite being the most treatable mental health disorder[2], it takes **11 years** on average to get treated[3].

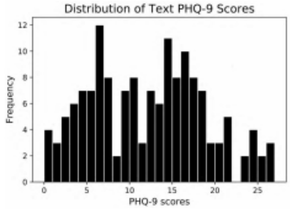
Suicide is the 2nd leading cause of death for US adults under 30. Globally depression is the leading cause of **disability**, costing $1 trillion[3].

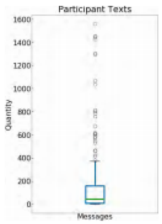
Given texting popularity, **text messages** could be used to passively screen for depression but only a **third** of people are willing to share this modality[4].

Data

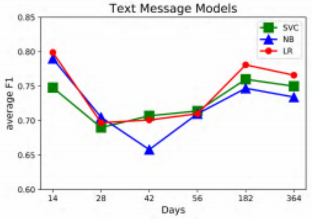
PHQ-9 score	Interpretation[2]	Treatment
0-4	Not Depressed	NA
5-9	Mildly Symptomatic	Monitor
10-14	Mild Depression	Support
15-19	Moderate Depression	Treatment
20+	Severe Depression	Treatment

Moodable[4]/EMU data: retrospectively-harvested crowd-sourced Smartphone & social media data. PHQ-9 was deployed to obtain a depression label. 151 participants sent texts within the last year[5].



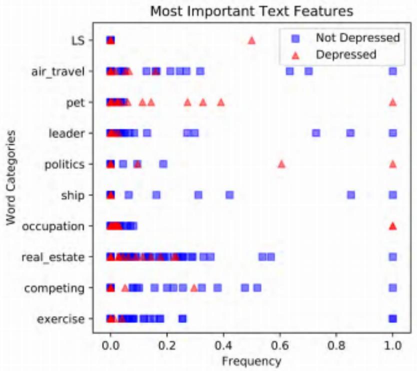


Screening with Text Messages

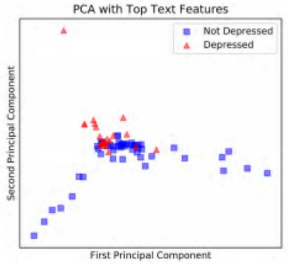


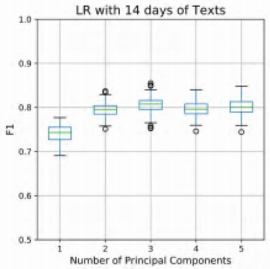
Machine learning methods selected from 245 content features involving:

- Word category frequencies
- POS tag frequencies
- Sentiment
- Volume



Logistic regression models only used 10 features from **two weeks** of texts, achieving an **F1 = 0.81** with three principal components[5].





Generating Text Messages

Goal: create a corpus of **public texts** from PHQ-9 labeled participants.

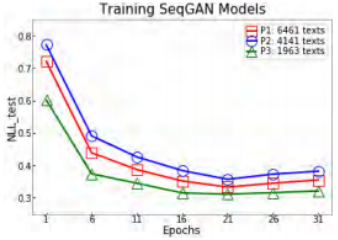
Generative Adversarial Networks (GANs) generate realistic data by using a **generator and a discriminator** engaged in a minimax game. GANs must be modified to generate sequences of discrete tokens[6] as

1. words are not differentiable leading to **no policy updates** and
2. sequences are only scored when complete so rewards are sparse.

Evolution of Text Generation Models

LSTM → SeqGAN → RankGAN → LeakGAN → RelGAN

We deploy **SeqGAN** to determine the impact of text quantity on generation quality measured by negative log-likelihood (NLL). SeqGAN
1. trains a stochastic parameterized policy with a policy gradient and
2. estimates rewards using a Monte Carlo search with a roll-out policy.



SeqGAN can still be **effective** when trained on around 2000 texts, though most of the participants have under 200 texts. We only need **20 epochs** to train.

Generated Text Message Examples

sure how much how awesome! ▪ *let me know when you see Monday*
aww they'll be like soon ▪ *sure sound fine so* ▪ *ok. i can come tonight*
actually kids were on this way home ▪ *should to make the toll on lol*

Future Work in Generating Texts

- Compare the screening ability of real texts with texts generated by GANs built on texts from **single and multiple participants**.
- Further anonymize generated texts by replacing named entities.
- Evaluate the appropriateness of popular metrics for this task.

References

[1] Evans, Bira, Gastelum, Weiss, Vanderford. "Evidence for a Mental Health Crisis in Graduate Education," Nature Biotechnology, 2018.
[2] Kroenke, Spitzer, William. "The PHQ-9: Validity of a Brief Depression Severity Measure," Journal of General Internal Medicine, vol. 16(9), 2001.
[3] National Alliance on Mental Health. "Mental Health By Numbers," 2019. Accessed 2020.
[4] Dogrucu, et al. "Instantaneous Depression Assessment using Machine Learning on Voice Samples and Retrospectively Harvested Smart-phone and Social Media Data," Smarthealth, accepted.
[5] Tlachac, Rundensteiner. "Screening for Depression with Retrospectively Harvested Private versus Public Text," IEEEjBHI, 2020.
[6] Yu, et al. SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient," AAAI, 2017.

Acknowledgments

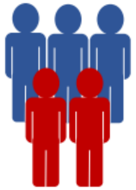
- US Department of Education P200A150306: GAANN Fellowships
- Ermal Toto, Nick Pingal, Samuel S. Ogden, Marissa Bennett, Francis Castro
- DSRG and DLRG communities
- Prof. Agu, Dogrucu, Peruic, Isaro, and Ball, Gao, Flannery, Resom, Assan, and Wu

Worcester Polytechnic Institute

Depression Screening with Text Messages

ML Tlachac, Data Science PhD Candidate   Advisor: Elke Rundensteiner

Motivation



2 in 5 graduate students suffer from **depression**[1].

Despite being the most treatable mental health disorder[2], it takes **11 years** on average to get treated[3].

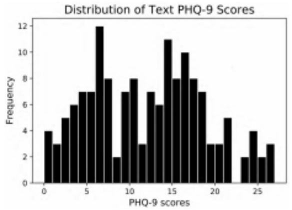
Suicide is the 2nd leading cause of death for US adults under 30. Globally depression is the leading cause of **disability**, costing $1 trillion[3].

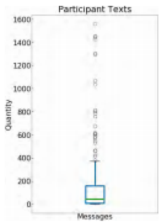
Given texting popularity, **text messages** could be used to passively screen for depression but only a **third** of people are willing to share this modality[4].

Data

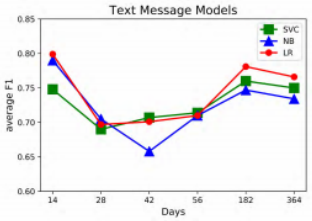
PHQ-9 score	Interpretation[2]	Treatment
0-4	Not Depressed	NA
5-9	Mildly Symptomatic	Monitor
10-14	Mild Depression	Support
15-19	Moderate Depression	Treatment
20+	Severe Depression	Treatment

Moodable[4]/EMU data: retrospectively-harvested crowd-sourced Smartphone & social media data. PHQ-9 was deployed to obtain a depression label. 151 participants sent texts within the last year[5].



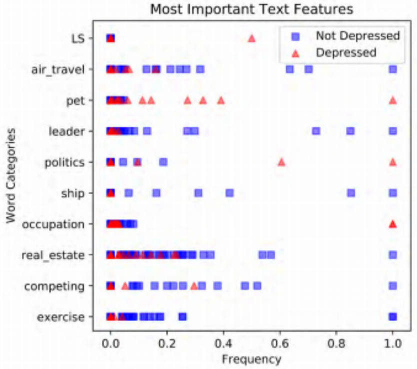


Screening with Text Messages

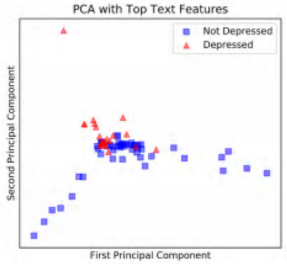


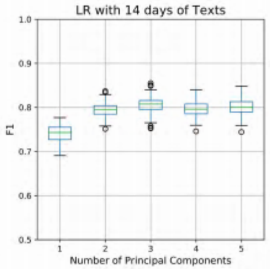
Machine learning methods selected from 245 content features involving:

- Word category frequencies
- POS tag frequencies
- Sentiment
- Volume



Logistic regression models only used 10 features from **two weeks** of texts, achieving an **F1 = 0.81** with three principal components[5].





Generating Text Messages

Goal: create a corpus of **public texts** from PHQ-9 labeled participants.

Generative Adversarial Networks (GANs) generate realistic data by using a **generator and a discriminator** engaged in a minimax game. GANs must be modified to generate sequences of discrete tokens[6] as

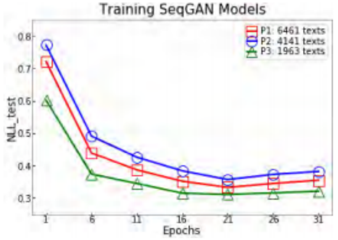
1. words are not differentiable leading to **no policy updates** and
2. sequences are only scored when complete so rewards are sparse.

Evolution of Text Generation Models

LSTM ➔ SeqGAN ➔ RankGAN ➔ LeakGAN ➔ RelGAN

We deploy **SeqGAN** to determine the impact of text quantity on generation quality measured by negative log-likelihood (NLL). SeqGAN

1. trains a stochastic parameterized policy with a policy gradient and
2. estimates rewards using a Monte Carlo search with a roll-out policy.



SeqGAN can still be **effective** when trained on around 2000 texts, though most of the participants have under 200 texts. We only need **20 epochs** to train.

Generated Text Message Examples

sure how much how awesome! • *let me know when you see Monday*
aww they'll be like soon • *sure sound fine so* • *ok. i can come tonight*
actually kids were on this way home • *should to make the toll on lol*

Future Work in Generating Texts

- Compare the screening ability of real texts with texts generated by GANs built on texts from **single and multiple participants**.
- Further anonymize generated texts by replacing named entities.
- Evaluate the appropriateness of popular metrics for this task.

References

[1] Evans, Bira, Gastelum, Weiss, Vanderford. "Evidence for a Mental Health Crisis in Graduate Education," Nature Biotechnology, 2018.
[2] Kroenke, Spitzer, William. "The PHQ-9: Validity of a Brief Depression Severity Measure," Journal of General Internal Medicine, vol. 16(9), 2001.
[3] National Alliance on Mental Health. "Mental Health By Numbers," 2019. Accessed 2020.
[4] Dogrucu, et al. "Instantaneous Depression Assessment using Machine Learning on Voice Samples and Retrospectively Harvested Smart-phone and Social Media Data," Smarthealth, accepted.
[5] Tlachac, Rundensteiner. "Screening for Depression with Retrospectively Harvested Private versus Public Text," IEEEjBHI, 2020.
[6] Yu, et al. SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient," AAAI, 2017.

Acknowledgments

- US Department of Education P200A150306: GAANN Fellowships
- Ermal Toto, Nick Pingal, Samuel S. Ogden, Marissa Bennett, Francis Castro
- DSRG and DLRG communities
- Prof. Agu, Dogrucu, Peruic, Isaro, and Ball, Gao, Flannery, Resom, Assan, and Wu

Year: 2020
Tool: PowerPoint

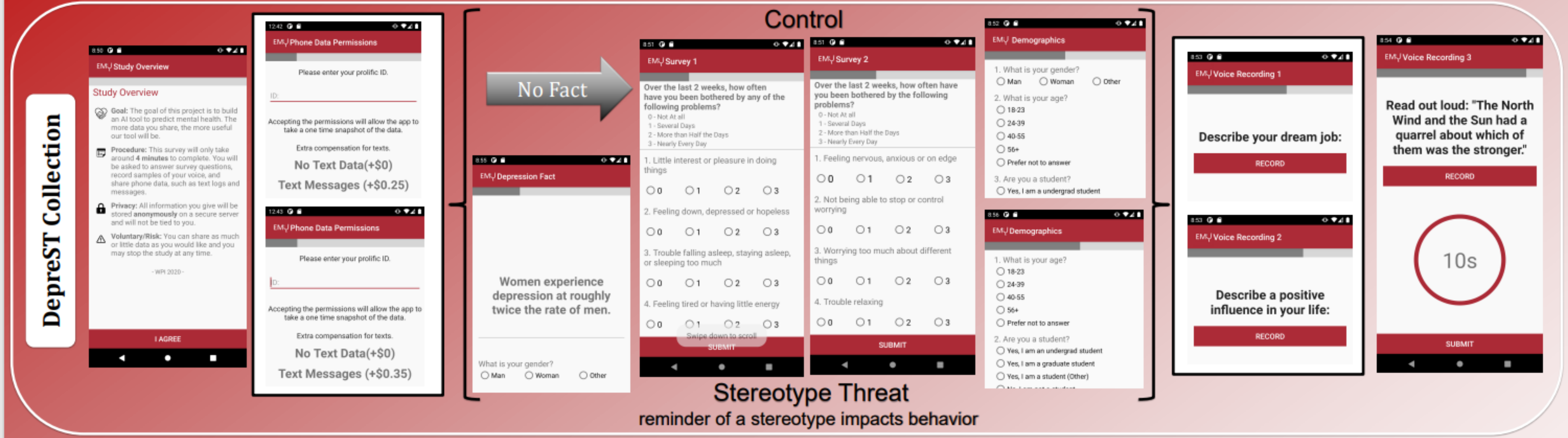
Self Critiques

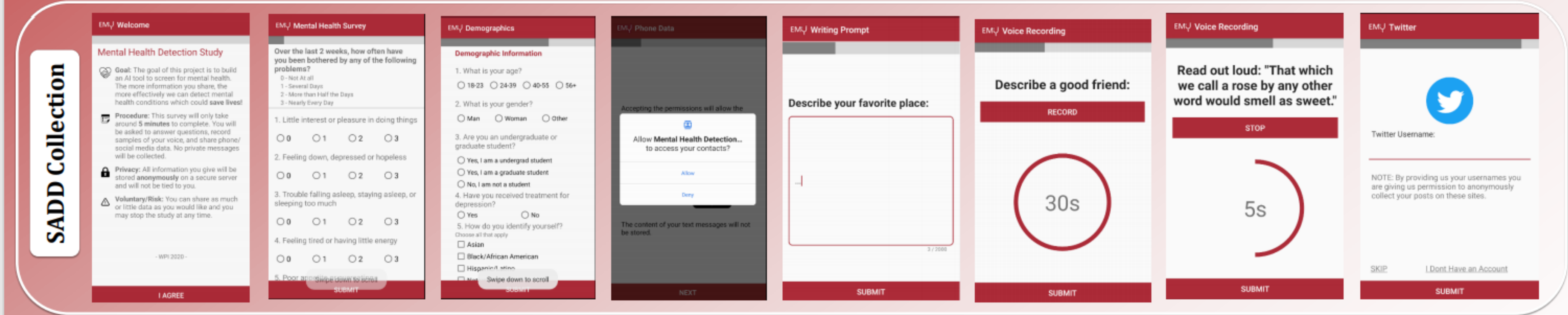
- Story too big for a poster
- Why mention future work?
- Important words in red
- Generated text examples are fun for readers

Mobile Data Collections for Mental Illness Screening

ML Tlachac, Data Science PhD Candidate
Advisor: Elke Rundensteiner

Year: 2021
Tool: PowerPoint



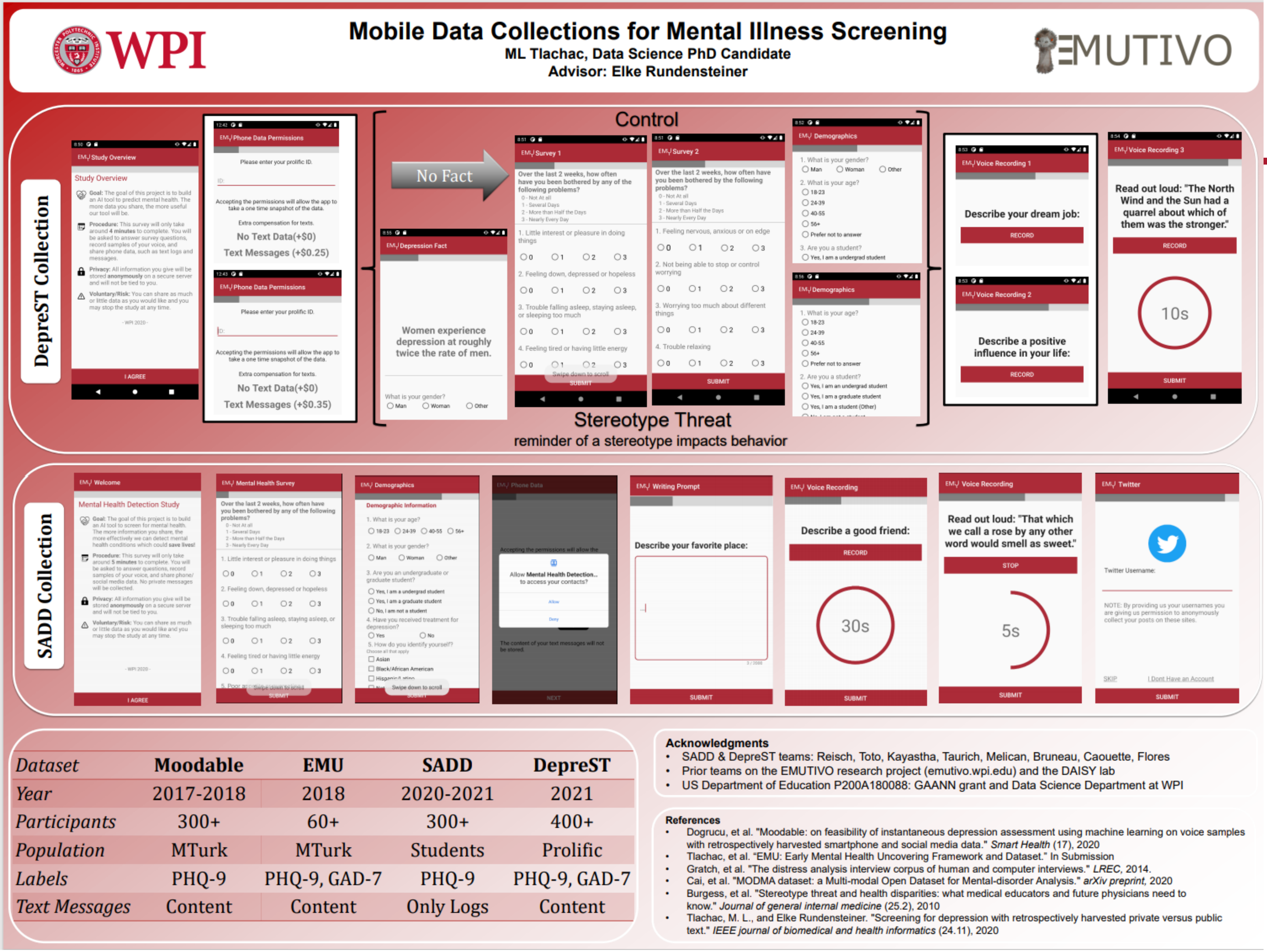


Dataset	Moodable	EMU	SADD	DepreST
Year	2017-2018	2018	2020-2021	2021
Participants	300+	60+	300+	400+
Population	MTurk	MTurk	Students	Prolific
Labels	PHQ-9	PHQ-9, GAD-7	PHQ-9	PHQ-9, GAD-7
Text Messages	Content	Content	Only Logs	Content

Acknowledgments
- SADD & DepreST teams: Reisch, Toto, Kayastha, Taurich, Melican, Bruneau, Caouette, Flores
- Prior teams on the EMUTIVO research project (emutivo.wpi.edu) and the DAISY lab
- US Department of Education P200A180088: GAANN grant and Data Science Department at WPI

References
- Dogrucu, et al. "Moodable: on feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data." *Smart Health* (17), 2020
- Tlachac, et al. "EMU: Early Mental Health Uncovering Framework and Dataset." In Submission
- Gratch, et al. "The distress analysis interview corpus of human and computer interviews." *LREC*, 2014.
- Cai, et al. "MODMA dataset: a Multi-modal Open Dataset for Mental-disorder Analysis." *arXiv preprint*, 2020
- Burgess, et al. "Stereotype threat and health disparities: what medical educators and future physicians need to know." *Journal of general internal medicine* (25.2), 2010
- Tlachac, M. L., and Elke Rundensteiner. "Screening for depression with retrospectively harvested private versus public text." *IEEE journal of biomedical and health informatics* (24.11), 2020

Worcester Polytechnic Institute



Year: 2021
Tool: PowerPoint

Self Critiques

- Can't read text in screenshots
- Can't match references
- No unnecessary words on poster
- Visually interesting

Worcester Polytechnic Institute

Mobile Depression Screening with Time Series of Text Logs and Call Logs

ML Tlachac, Veronica Melican, Miranda Reisch, Elke Rundensteiner
Worcester Polytechnic Institute
4-page paper @ IEEE BHI-BSN 2021

Year: 2021
Tool: PowerPoint

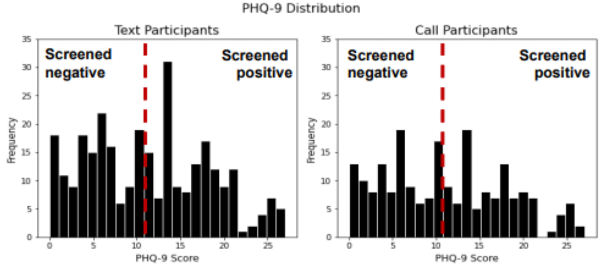
Research Questions

Given logs, is it best to screen for depression with:

1. Text logs or call logs?
2. Incoming, outgoing, or all communications?
3. Communication count, average length, or contacts?
4. Aggregation intervals of 4, 6, 12 or 24 hours?
5. Time series or features from time series?

The Data

Two weeks of logs from the Moodable[1] and EMU[2] datasets labeled with PHQ-9 depression screening scores[3]. If PHQ-9 ≥ 10, screen positive for depression.

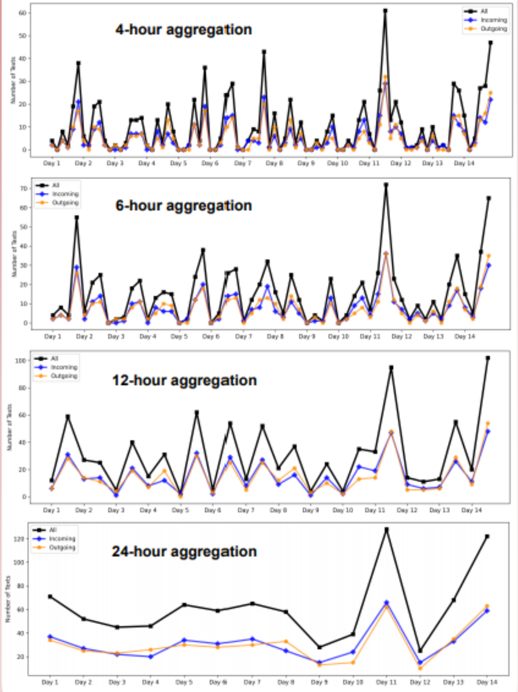


The 312 participants shared different types of logs:

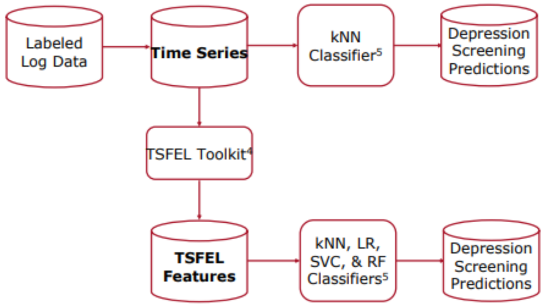
Log	All	Incoming	Outgoing
Text logs	245	290	99
Call logs	212	182	197

Creating the Time Series

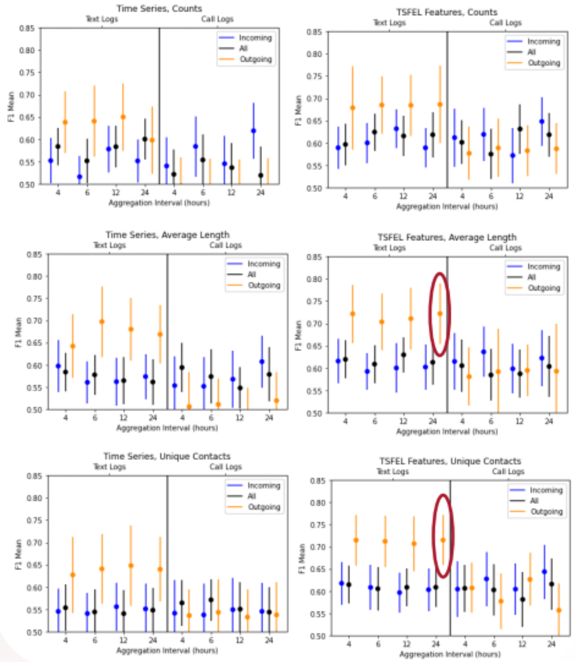
We create 72 sets of time series from the logs using the communication count, average length of communications, and number of contacts. We aggregate these values every 4, 6, 12, & 24 hours. All text log time series for a single participant:



Machine Learning Pipeline



Screening Results



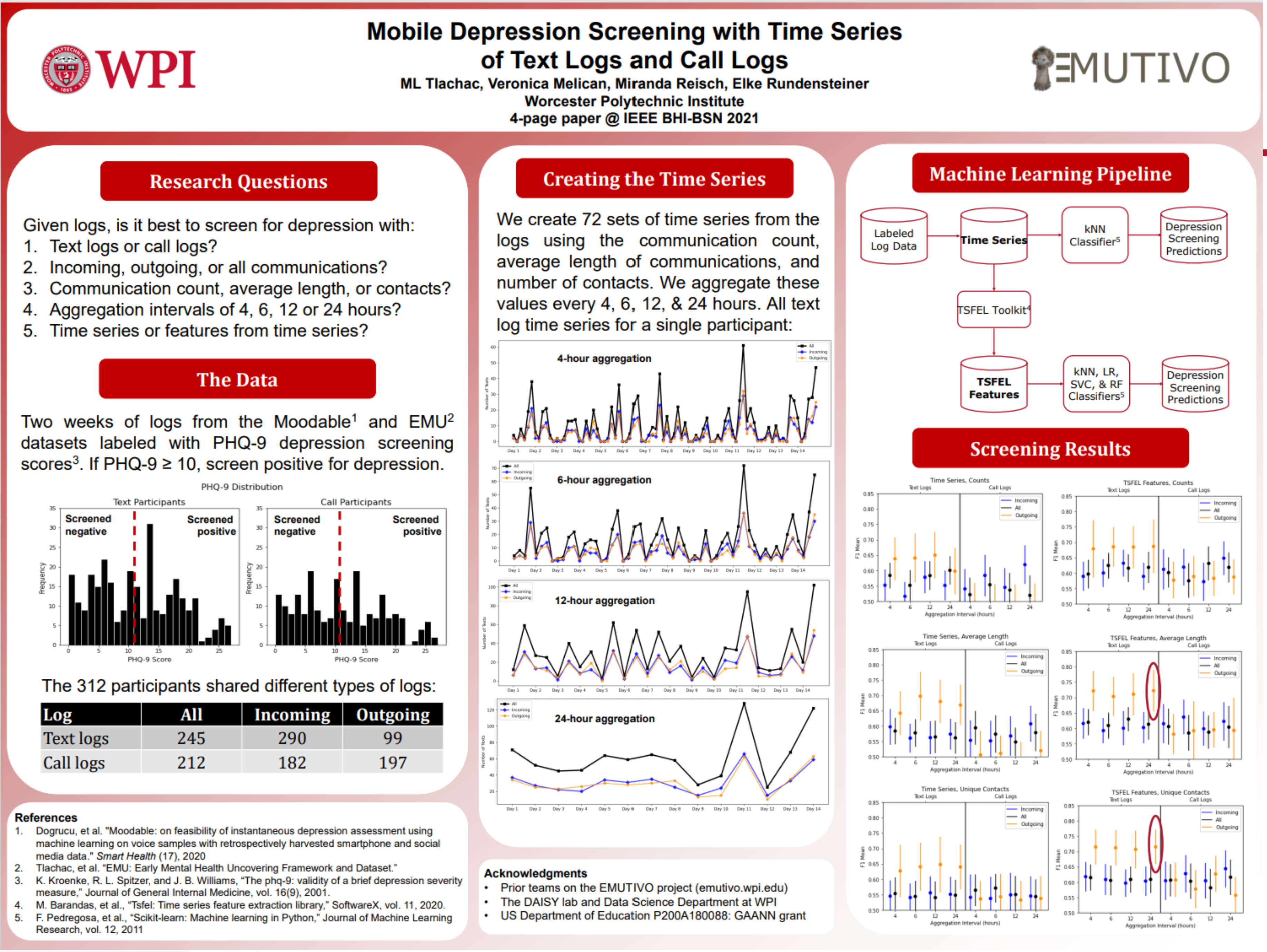
References

1. Dogrucu, et al. "Moodable: on feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data." *Smart Health* (17), 2020
2. Tlachac, et al. "EMU: Early Mental Health Uncovering Framework and Dataset."
3. K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The phq-9: validity of a brief depression severity measure," Journal of General Internal Medicine, vol. 16(9), 2001.
4. M. Barandas, et al., "Tsfel: Time series feature extraction library," SoftwareX, vol. 11, 2020.
5. F. Pedregosa, et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, 2011

Acknowledgements

- Prior teams on the EMUTIVO project (emutivo.wpi.edu)
- The DAISY lab and Data Science Department at WPI
- US Department of Education P200A180088: GAANN grant

Worcester Polytechnic Institute



Year: 2021
Tool: PowerPoint

Self Critiques

- Colors are unbalanced
- Title of result plots are too small to read
- Large font
- Fun pipeline

Worcester Polytechnic Institute

Poster Tips

All Posters

- Know your audience
- Minimize words
  - Text big enough to read
  - Readable text/background colors
  - Consistent accessible font
- Maximize visualizations/tables
  - High quality images
- Make reading order clear
- Have an elevator pitch ready

Printed Posters for In-Person

- Use a column-based format
- Put while panels over colorful background to reduce ink
- Include a small white border in case printing misalignment
- Have notebook to write suggestions and/or contact info