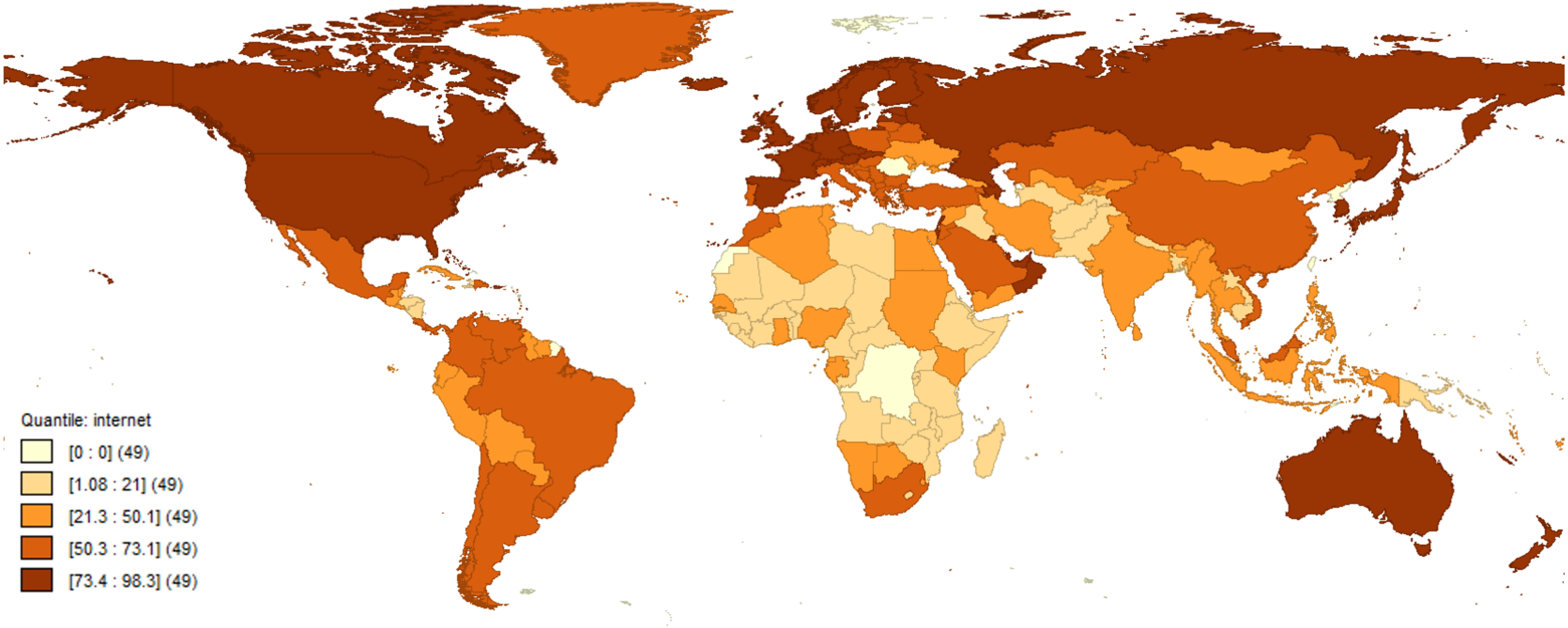
Bad Neighbors and The Internet

A GEOSPATIAL ANALYSIS OF INTERNET ADOPTION

Research Question

- Does geographic location matter in a country’s internet adoption?
- Are there spatial clusters (of countries) of low or high internet adoption?



2015 Quintile World Map of Internet Adoption

Our Contribution

- There is a rich literature dealing with the determinants of the internet adoption and the digital divide
- Little attention has been paid on the possibility of cross-country (spatial) interactions in the adoption process

The Importance of the Internet

Digitalization

- rewriting the rules of international competition
- bringing about many opportunities for newcomers to enter the global value chains and catch up with incumbents
- applicable not only to companies, but also countries

However, it is unclear:

- the digital revolution will help developing countries better integrate into the world economy
- or enable rich countries to sustain or accelerate their competitive advantage.

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Developed Countries	53.45	59.03	61.27	62.90	66.51	67.68	73.84	76.87	79.50	82.25
Developing Countries	9.40	11.92	14.64	17.42	21.07	24.05	26.99	29.51	32.41	35.28
Digital Divide	44.05	47.11	46.63	45.48	45.44	43.63	46.85	47.36	47.09	46.97

Thus, it will be important to look at the determinants of internet adoption in developing countries and relook at strategies to abridge the digital gap.

Data

Compiled panel data on 63 developing countries on telecommunication and technology usage, country demographic and institutional characteristics from 2000 to 2015.

Variable	Obs	Mean	Std. Dev.	Min	Max
Internet User	557	8.78348	9.966485	0.065239	56
Neighbor	472	12.17857	11.72404	0.255	57
Percentage of urban population	567	36.80471	14.61583	9.375	69.274
Telephone fixed lines	554	4.852888	6.240416	0	30.64515
GDP per capita	558	3261.009	2320.175	530.9611	10580.9
Average years of schooling	562	5.055872	2.643623	1.3	12.1
Freedom of press	567	58.903	16.7821	24	97
Older than 64	567	4.262633	2.472458	2.176046	16.13981

Methodology

1. Adopt a spatial econometric methodology
2. Spatial dependence can be introduced in the regression specification in two ways
   1. Observational units (Spatial autocorrelation)
   2. Error term (Spatial error autocorrelation)
3. Conduct the diagnostic test for the presence of spatial autocorrelation and spatial error autocorrelation to determine model specification
4. Perform non-spatial diagnostic tests for multicollinearity, non-normality and heteroskedasticity
5. Perform Hausman Wu to determine the use of Random Effects and Fixed Effects estimation
6. Control for potential endogeneity problems with spatial lag of the dependent variable
   - most appropriate instruments are the spatially lagged explanatory variables

Methodology

According to the results of spatial diagnostic tests, our model is specified with a spatial lag and control variables:

$$internet_{i,t} = \alpha + Winternet_{i,t} + \beta^{'}X + \alpha_i + \varepsilon_i$$

- Dependent variable: Percentage of the population as internet users in a country
- α: Constant for all observations
- X: vector of control variables
- $Winternet_{i,t}$: spatial lag of dependent variable
- α_i: unobserved heterogeneity for each country
- ε_i: error term

Discovering and Predicting Video Game Experiential Genres

MACS 30200 Project

Chih-Yu Chiang

What are video game genres based on experiences in the games?

With the new genres,
how to understand video games better?

Experience and Video Game

Experience

One feels or is affected by to feel.

Video Game as an Experiential Product

Products dominantly emphasize on the consumption experiences.

- Cooper-Martin, E. (1991). *Consumers and movies: Some findings on experiential products*. NA-Advances in Consumer Research Volume 18.

Playing Video Games

A life projection that different players are motivated to seek out in the form of in-game experiences.

- Bartle, R. (1996). *Hearts, clubs, diamonds, spades: Players who suit MUDs*. The Journal of Virtual Environments, 1.
- Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). T*he motivational pull of video games: A self-determination theory approach*. Motivation and Emotion, 30(4), 347–363.

Why Bother?

Form		Experience
Shooter		Discover new world
Strategy		Unfold storyline
RPG		Collect virtual items
Action	VS	Experience a real war
Adventure		Vehicle Racing
Fighter		Destroy a city
Puzzle		Coop with teammates
Card		Lead a squad

All First-person Shooter Games

Similar Experience?

Coop with teammates

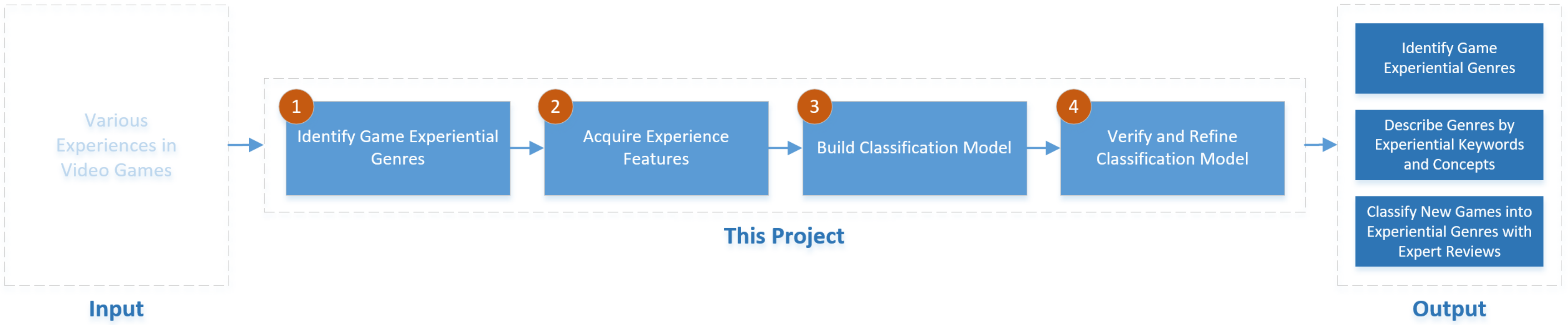
Experience unreal SWAT operations

Unfold storyline

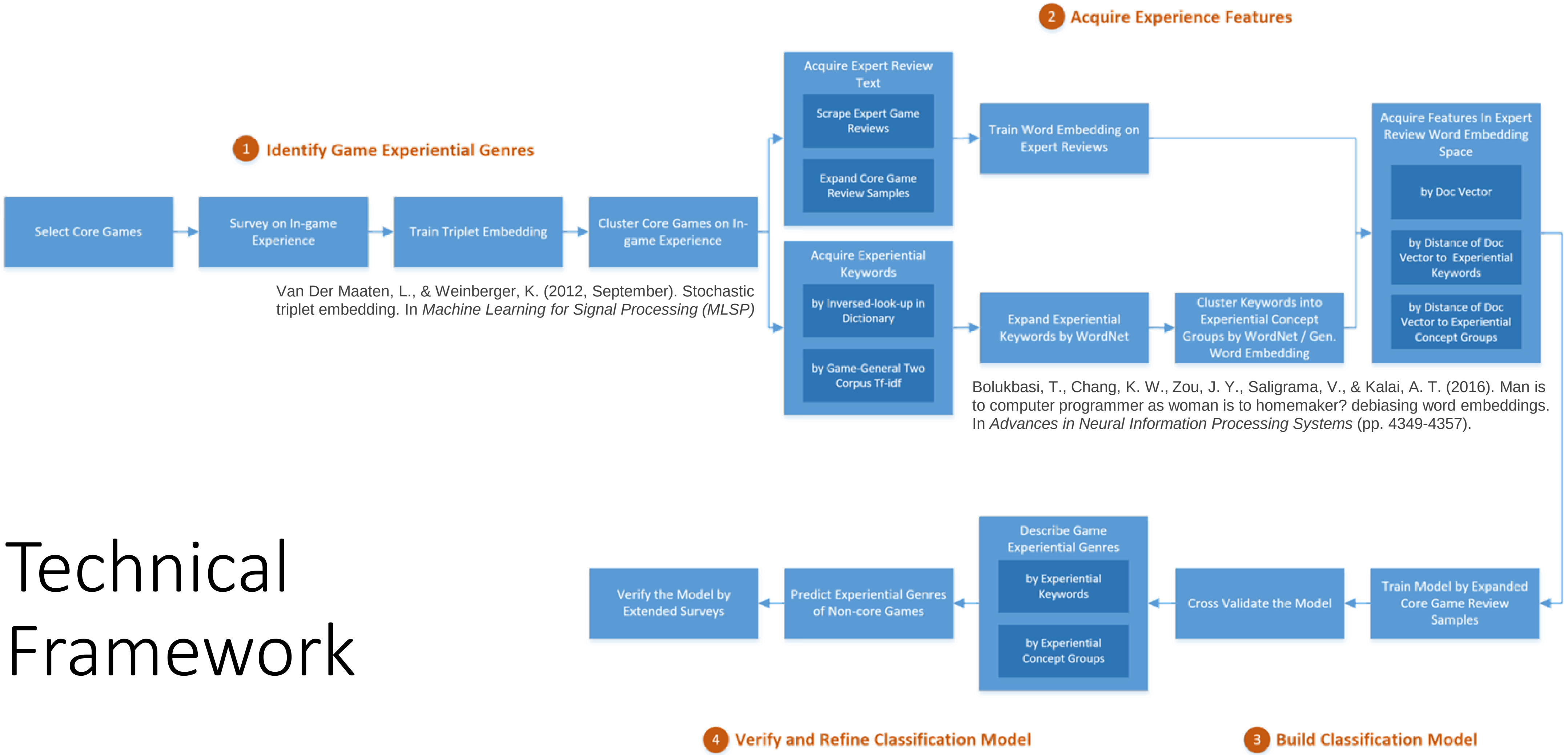
Coop with teammates

Experience real war

Project Overview



Technical Framework

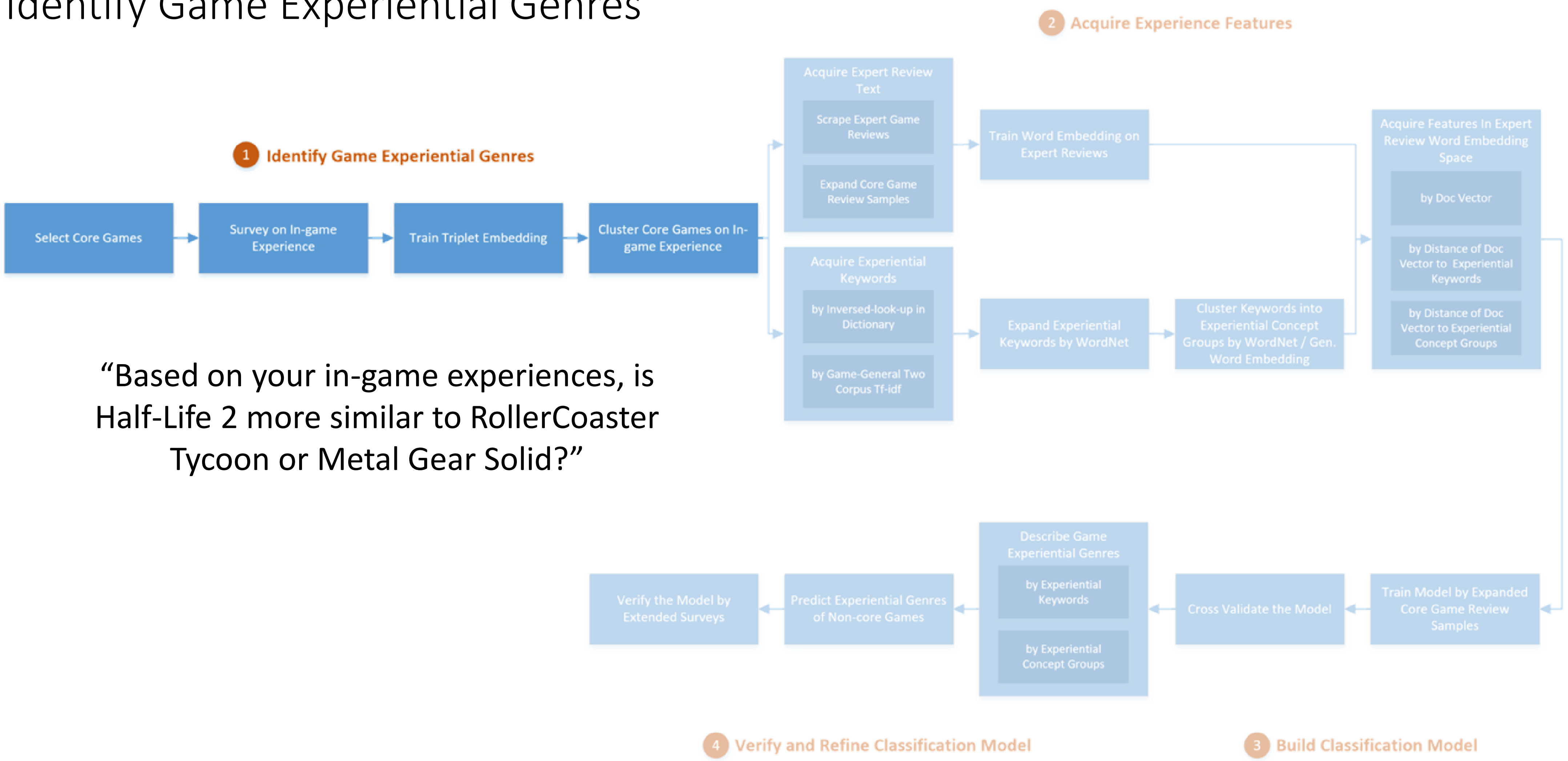


Van Der Maaten, L., & Weinberger, K. (2012, September). Stochastic triplet embedding. In *Machine Learning for Signal Processing (MLSP)*

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems* (pp. 4349-4357).

Technical Framework - 1

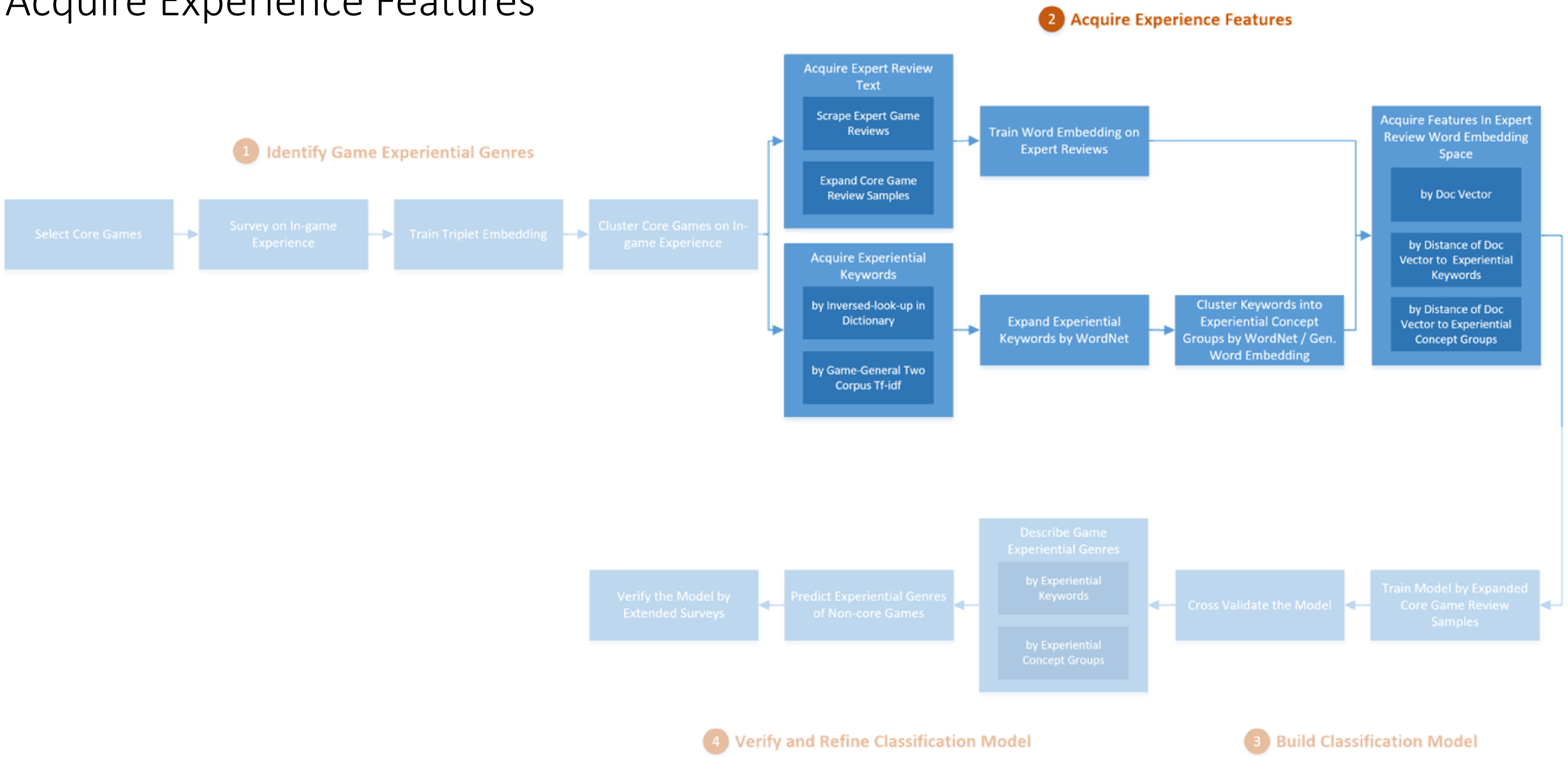
Identify Game Experiential Genres



"Based on your in-game experiences, is Half-Life 2 more similar to RollerCoaster Tycoon or Metal Gear Solid?"

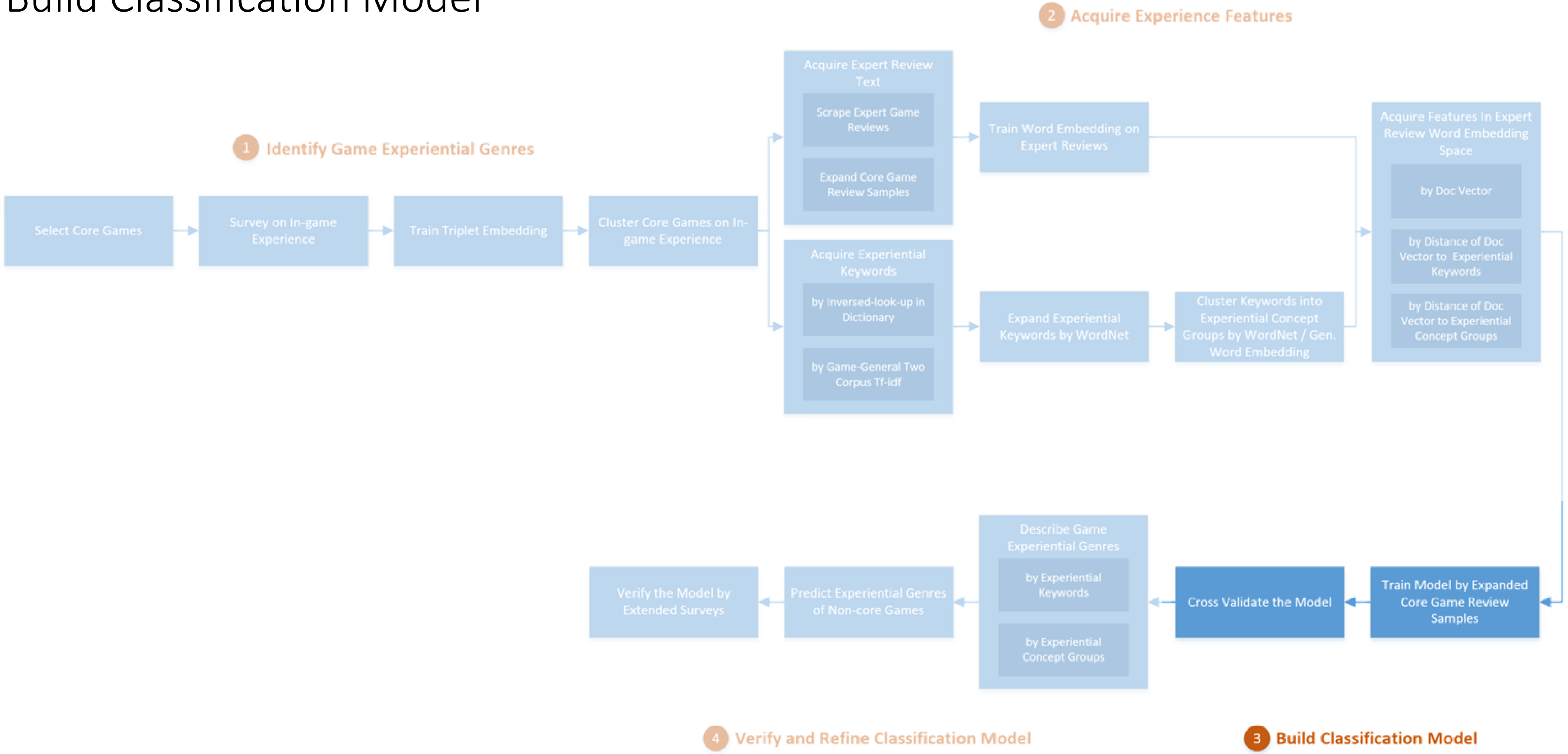
Technical Framework - 2

Acquire Experience Features



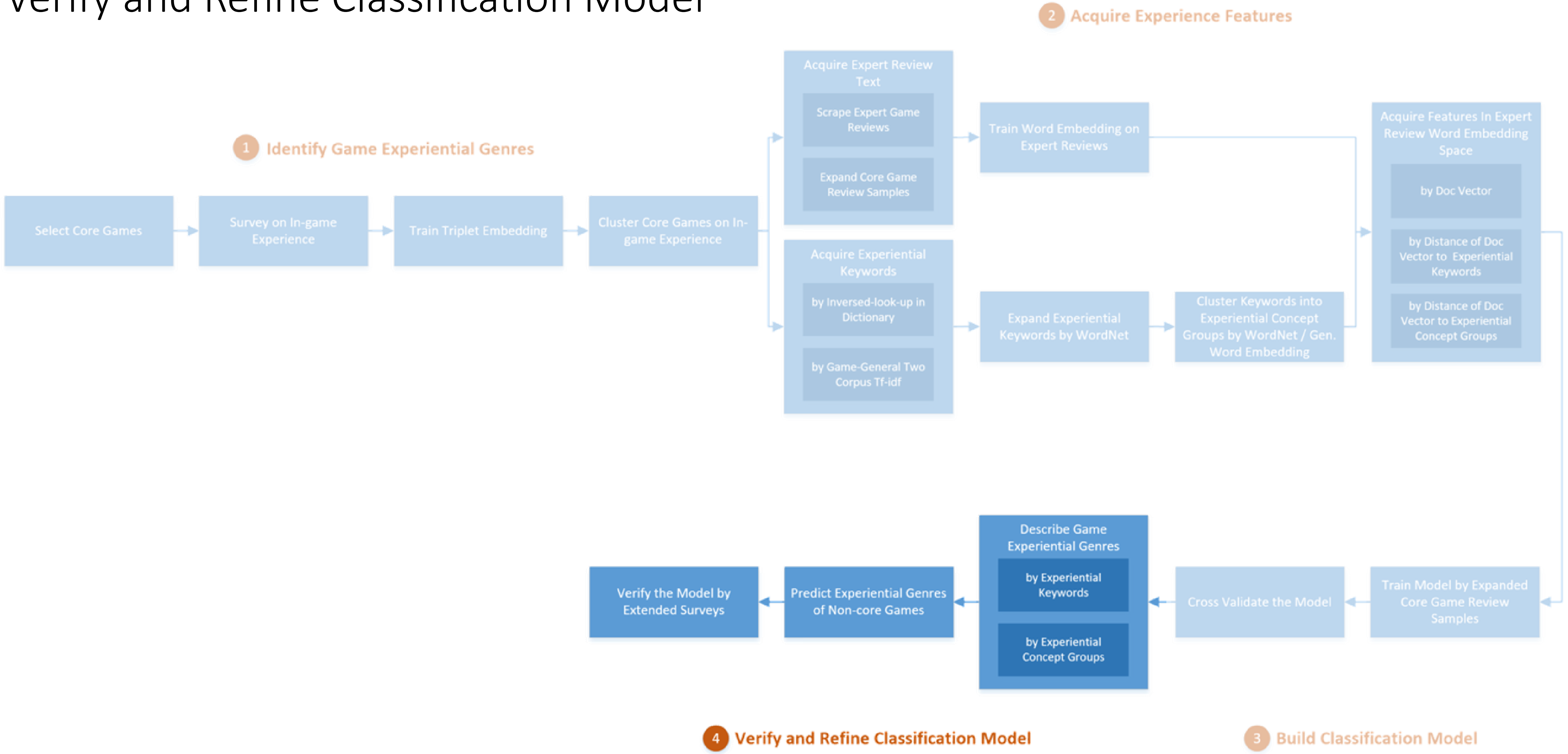
Technical Framework - 3

Build Classification Model

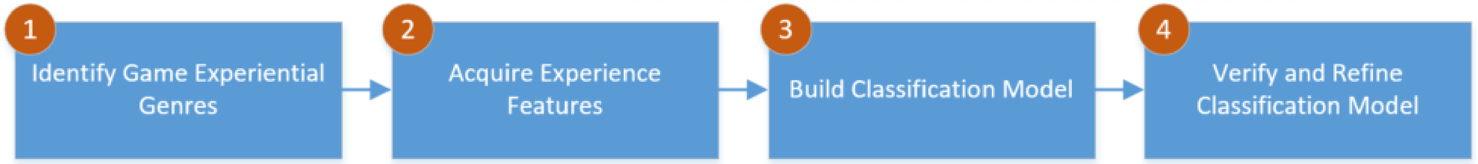


Technical Framework - 4

Verify and Refine Classification Model



Data

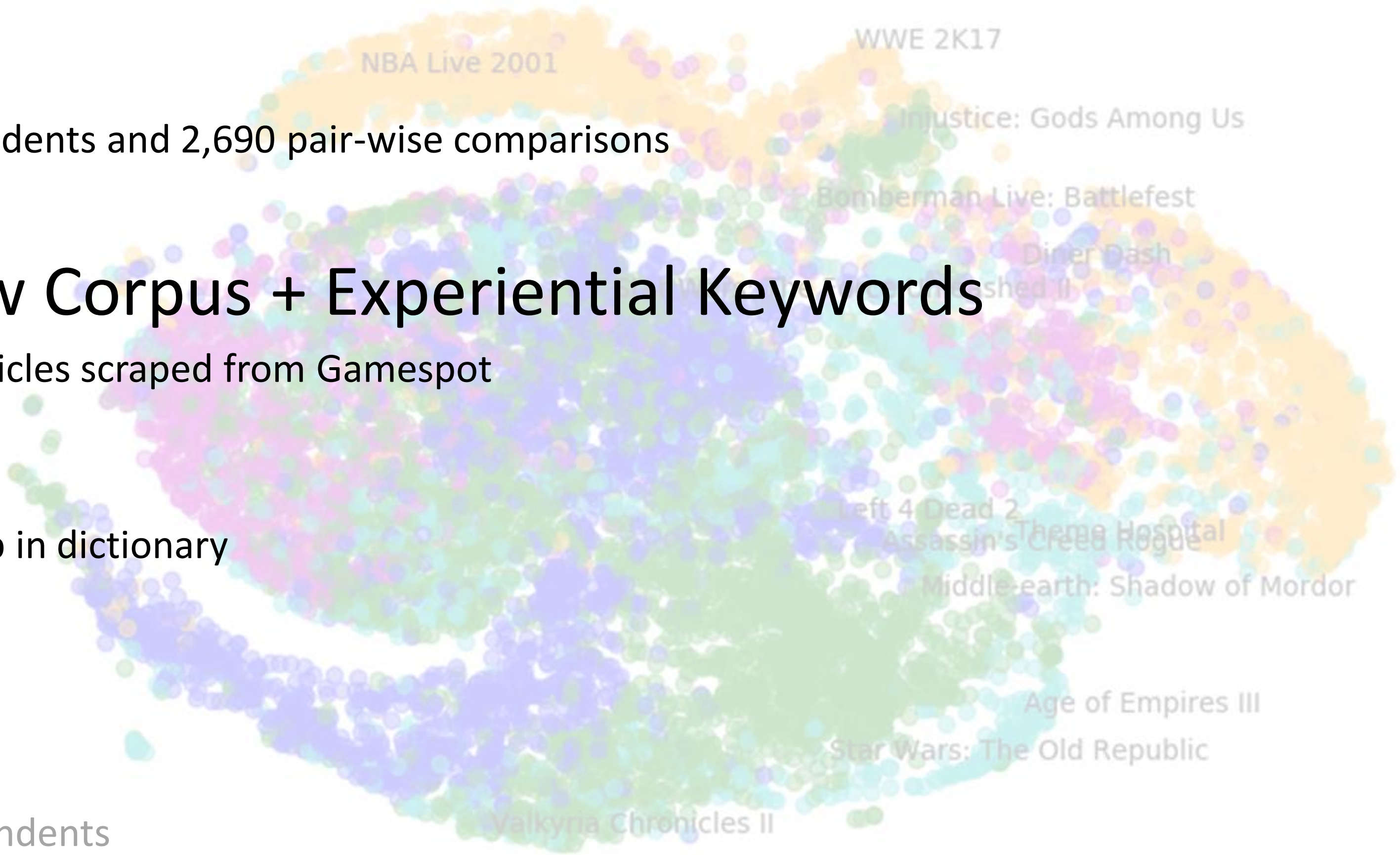


Stage 1: Triplet Survey

- Hand-picked 25 core games
- MTurk triplet survey with 135 respondents and 2,690 pair-wise comparisons

Stage 2 & 3: Expert Review Corpus + Experiential Keywords

- 11,022 expert video game review articles scraped from Gamespot
- 16,643,915 words
- 1,510 words per review on average
- 1,168 keywords from reverse look-up in dictionary



Stage 4: Validation Survey

- Expand to 40 core games
- Mturk survey with around 200 respondents

Result & Application

1. Identify Video Game Experiential Genres
2. Describe Genres by Experiential Keywords and Concepts
3. Classify New Games into Experiential Genres with Reviews

For Video Game Designers

As a research tool for understanding products and creating better in-game experience

For Video Game Players and Publishers

As a recommendation system based on similarity of in-game experience

For Movies, Music, Novels, and Other Experiential Products

Same applications!

Geographic Climate Regions and Homicide in the Continental US

A Challenge to Subcultural Arguments

Erin M. Ochoa

Subcultural Influence?

* Higher homicide rates in historical South
  * "Southern subculture of violence"
    * What this sometimes means: *Black* subculture of violence
* Messner (1983):
  * Higher homicide rates in cities in the Confederate South
  * Percent Black positively affects homicide rate for cities *outside* the South, but not those *within* it
* Peterson & Krivo (2010):
  * Controlling for neighborhood disadvantage *and* disadvantage in neighboring neighborhoods:
    * Gap in violent crime rates between Black & White neighborhoods shrinks dramatically (disappears when controlling for distal crime rate)

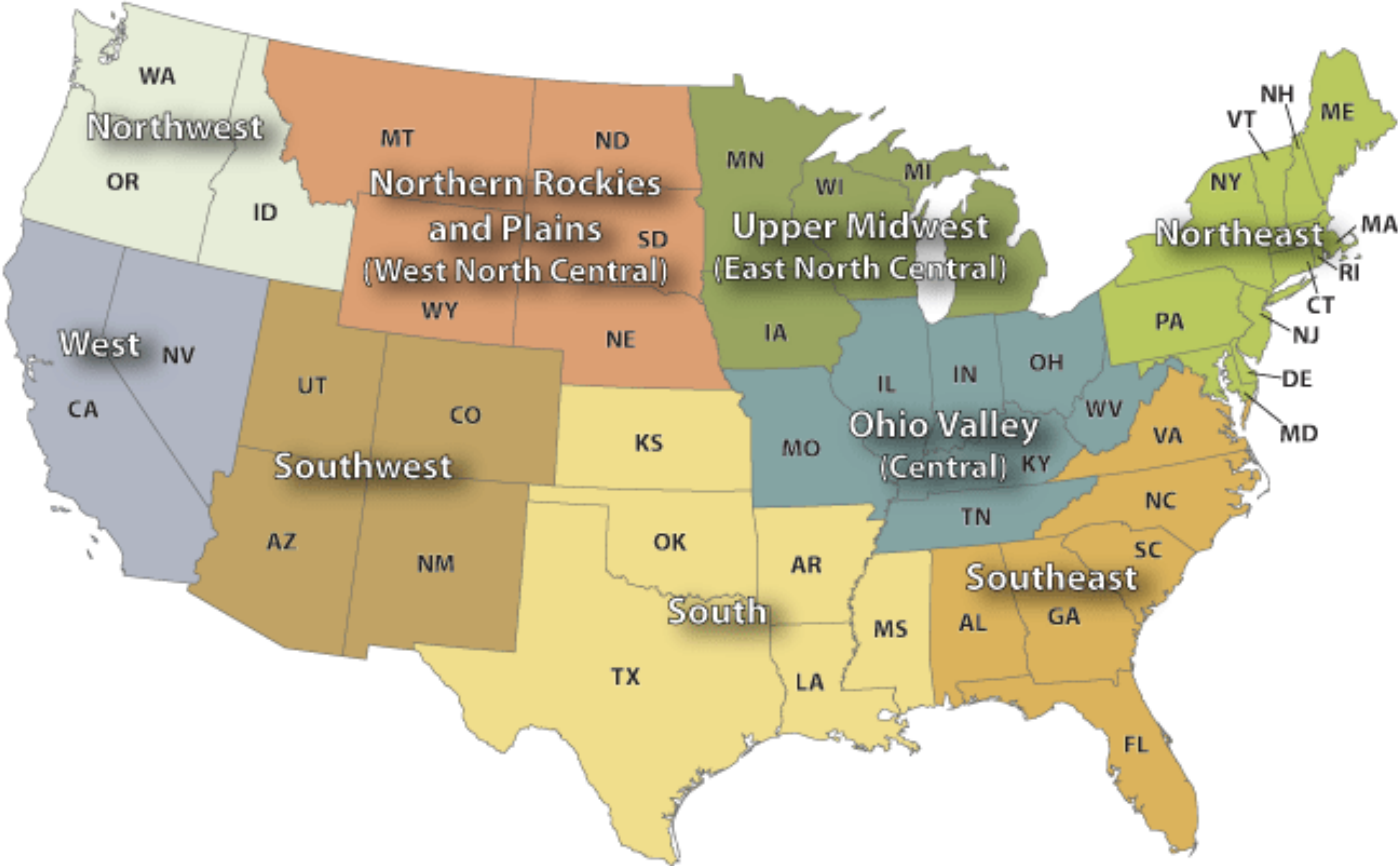
Research question:

What is the effect of climate on homicide rate?

Effect of Climate on Crime?

* Temperature affects violence through two general pathways (under a deterministic model):
    * Routine activities: people spend more time outdoors in pleasant weather (Rotton & Cohn, 2003)
    * Psycho-physiology: aggression increases in extreme conditions (see Anderson, 1989, for a review)
* Extend temperature to climate by including: dew point, cloudiness, precipitation, wind chill, and daylight hours

U.S. Climate Regions



NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Classification Approach: Statistical Learning

* Meteorological data (NOAA, 1981–2010)
* Daylight data (TSA, 1981–2010)
* Standard metropolitan statistical areas (SMSAs) in the continental U.S.
* Support vector machines (SVMs)
* Train a model classifying SMSAs by geographic climate region: attempt to match known regions

Controls: Demographics

* An index created from Census data:
  * Total population
  * Black population
  * Segregation (index of dissimilarity)
  * Young men (aged 15–34)
  * Low-wage workers
  * Low-prestige workers
  * Manufacturing jobs
  * Poor households
  * Female-headed households
  * Joblessness
  * No bachelor's degree
  * Residential instability
  * Residential loans
  * Immigrant prevalence

Dependent Variable

* Homicide rate (2008–2012) per 100,000 per SMSA
* Source: UCR monthly reports (ICPSR)

Geographic Climate versus Subculture

* Develop linear models with controls (dummy variables) for geographic climate regions and include climate measures:
    * Regions statistically significant: culture (and other environmental and political variables)
    * Regions not statistically significant: climate

References

* Craig A. Anderson (1989): "Temperature and Aggression: Ubiquitous Effects of Heat on Occurrence of Human Violence." *Psychological Bulletin* 106:1 (74–96).

* Steven F. Messner (1983): "Regional & Racial Effects on the Urban Homicide Rate: The Subculture of Violence Revisited." *American Journal of Sociology* 88:5 (997–1007).

* National Oceanic & Atmospheric Administration: "U.S. Climate Regions." https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php

* Ruth D. Peterson & Lauren J. Krivo (2010): *Divergent Social Worlds: Neighborhood Crime and the Racial-Spatial Divide*. New York: Sage.

* James Rotton & Ellen G. Cohn (2003): "Global Warming and U.S. Crime Rates: An Application of Routine Activity Theory." *Environment & Behavior* 35:6 (802–825).

Evaluating Bike Share as A Solution to the “Last Mile Problem” in Public Transit

Course: MACS 30200-Perspectives on Computational Research
Presenter: Bobae Kang
April 5, 2017

Agenda

- Background
- Research Question
- Data
- Model
- Summary

Background: bike share

- First introduced in the 1960s; became popular in the 2010s
- Historical development (DeMaio, 2009; Shaheen, Guzman & Zhang, 2010)
  - 1st gen: simply providing free bikes; no accountability
  - 2nd gen: docking stations, coin-deposit system; still little accountability
  - 3rd gen: incorporating information tech; preventing bike theft successfully
- Few studies quantitatively evaluating bike share usage (Faghih-Imani & Eluru, 2015)
- A complement or substitute for the existing public transit?

Background: last mile problem

- “Transit that offers frequent and rapid service along the main lines but leaves the travelers a mile from their destinations with poor connecting options is rarely the mode of choice” (Zellner, Massey, Shiftan, Levine, & Arquero, 2016).
- Bike share is often seen as a solution to this last mile problem
  - Some survey studies provide evidence for this statement (e.g., Martin & Shaheen, 2014)
  - Few studies have evaluated the statement by examining actual bike share use

Background: Divvy

- First launched in summer 2013
- >6,000 bikes, >580 docking stations
- Annual membership and 24-hour pass
- A Chicago Department of Transportation (CDOT) program
- Per-trip data available at Divvy webpage (www.divvybikes.com/system-data)

Image source: City of Chicago official website.

Research Question

- Does bike share serve as a solution to the last mile problem in public transportation?
- For all trips made from and to Divvy stations in proximity with CTA stops, is the likelihood of trips that are potentially multimodal i.e., made in connection with the public transit rides, greater than by random chance? If so, to what extent?

Data

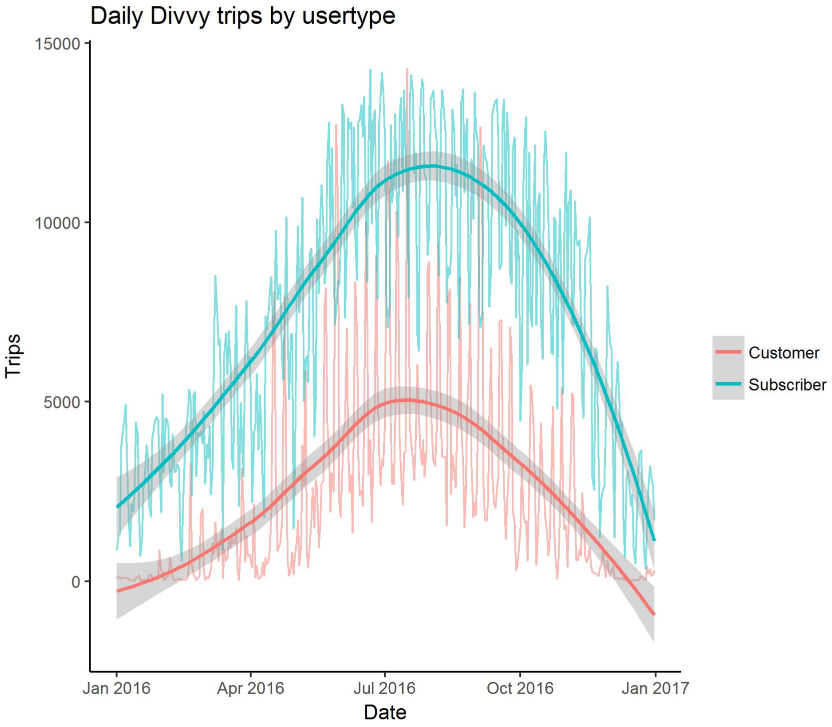
- Divvy trip data for the year of 2016
  - 3.6 million individual bike trips
  - Start and end time, start and end location, trip duration, user type, gender, birth year
  - 581 docking stations and their locations (longitude and latitude)
- CTA data
  - All CTA stops and their locations (11,487 stops)
  - Scheduled time for all bus routes and rail lines for all stops (2.8 million observations)
- Others
  - Weather in Chicago (temperature and precipitation)
  - Other geospatial features (CBD, community area, demographic data, etc.)

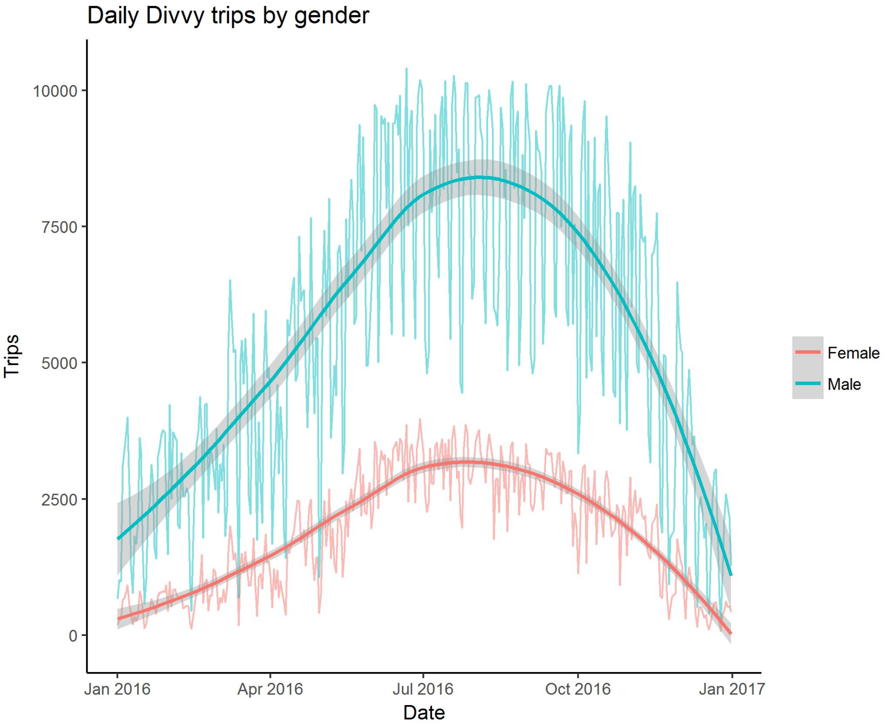
Data: Divvy trips

Table 1: Descriptive summary of Divvy trips

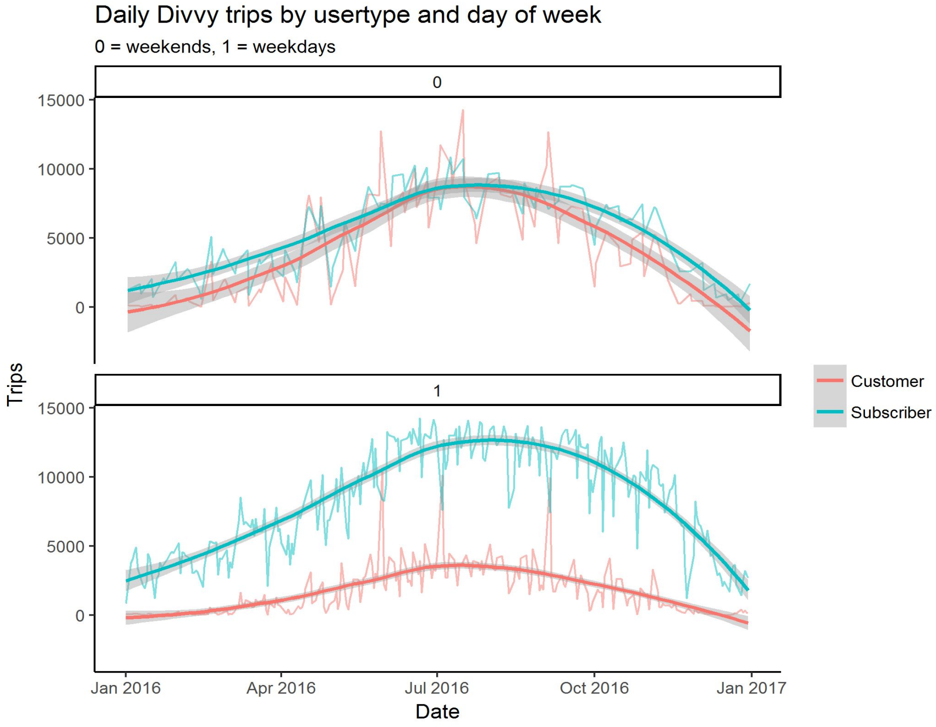
Trip attributes	All Users		Members		Daily Customers	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Annual membership	76.1%	-	-	-	-	-
Gender	-	-	74.8%	-	-	-
Age	-	-	35.52	10.75	-	-
Weekday	72.5%	-	79.6%	-	49.6%	-
Rush hour	44.5%	-	52.7%	-	18.4%	-
Duration (min)	16.56	31.54	12.04	20.76	30.96	50.18
Duration (male)	-	-	11.57	19.87	-	-
Duration (female)	-	-	13.44	23.12	-	-
Proximity, from (50m)	46.8%	-	48.3%	-	42.0%	-
Proximity, from (100m)	72.5%	-	74.4%	-	66.5%	-
Proximity, from (200m)	88.0%	-	90.3%	-	80.9%	-
Proximity, from (300m)	92.5%	-	94.3%	-	87.1%	-
Proximity, to (50m)	47.0%	-	48.3%	-	42.9%	-
Proximity, to (100m)	72.6%	-	74.4%	-	66.8%	-
Proximity, to (200m)	88.2%	-	90.4%	-	81.1%	-
Proximity, to (300m)	92.6%	-	94.3%	-	86.9%	-

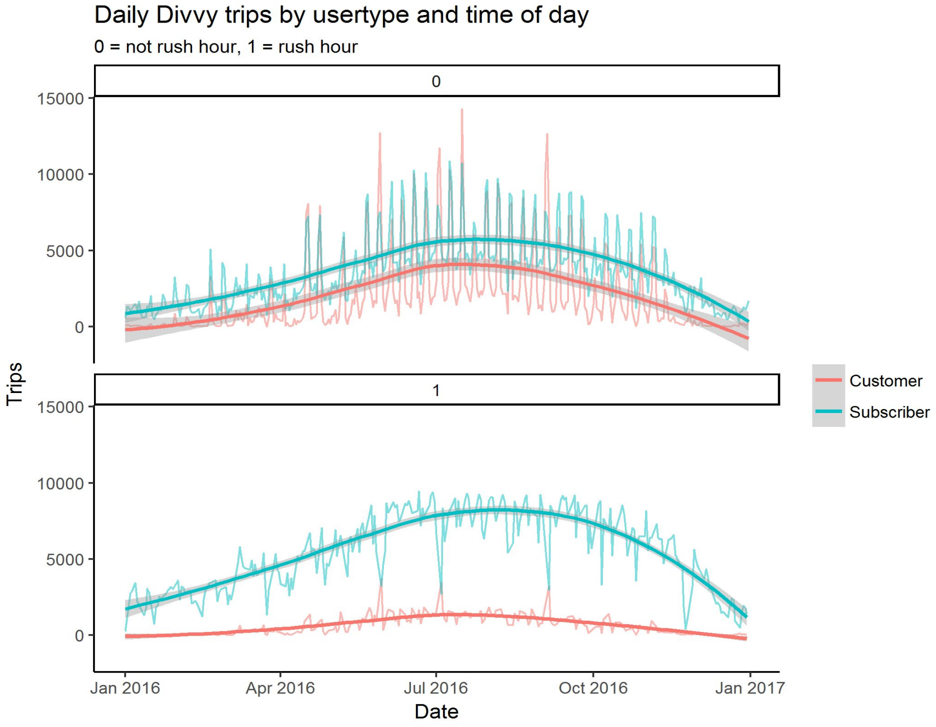
Data: Divvy trips



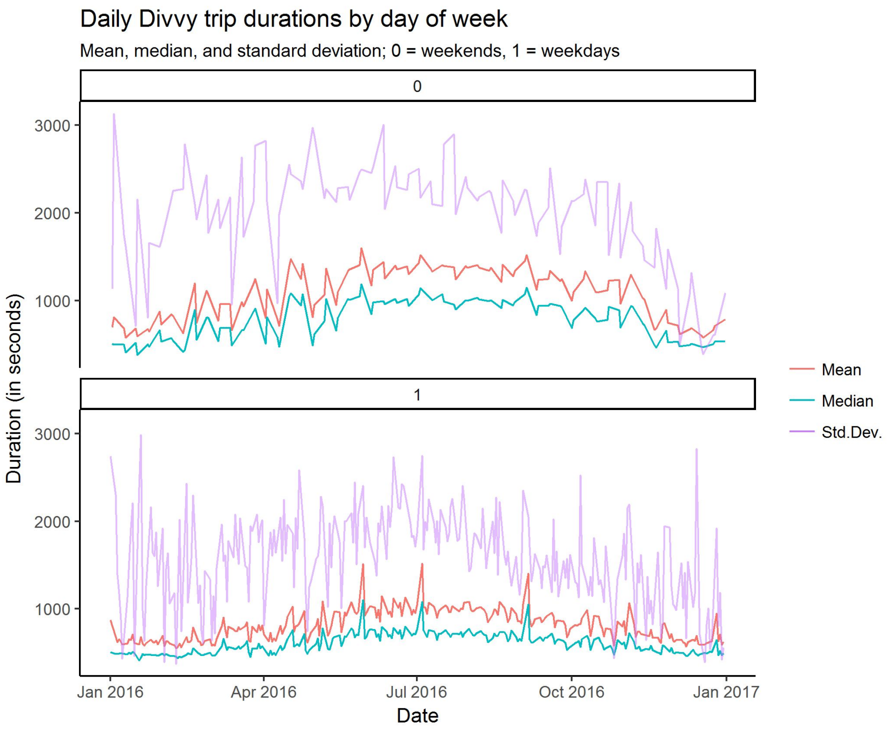


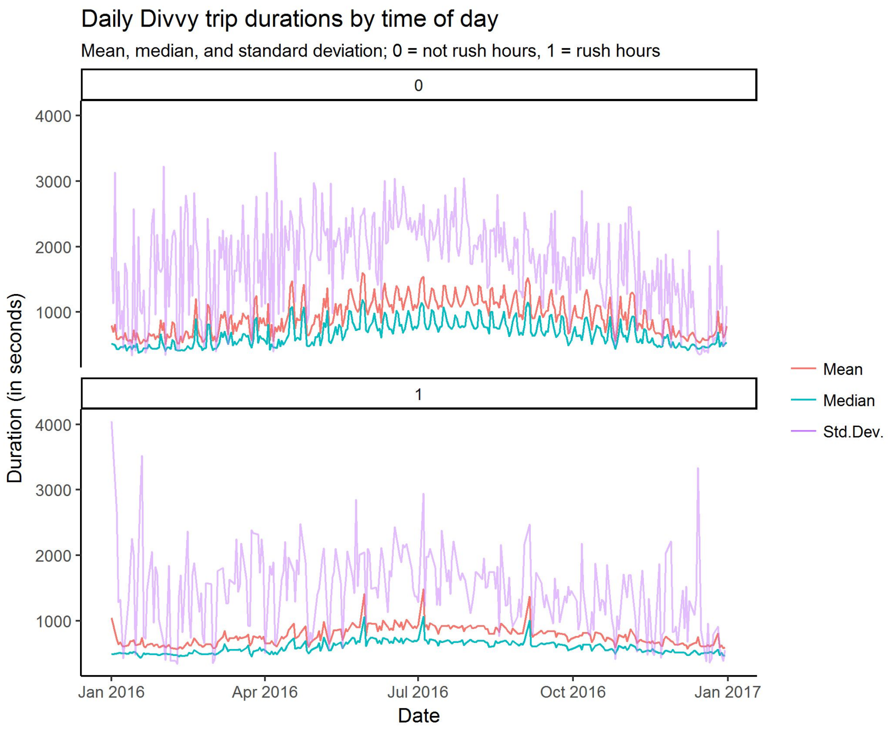
Data: Divvy trips





Data: Divvy trips

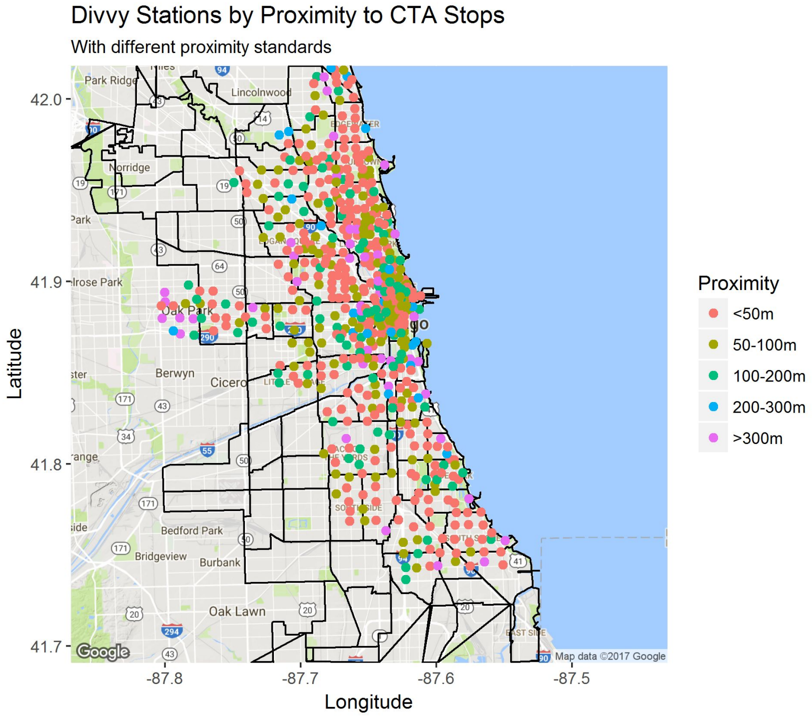




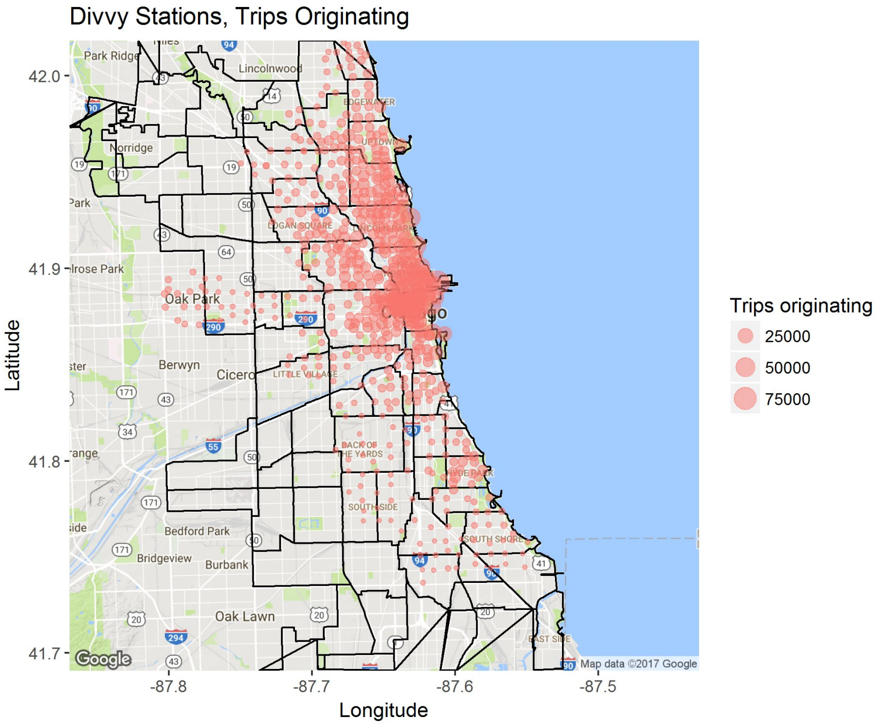
Data: Divvy stations

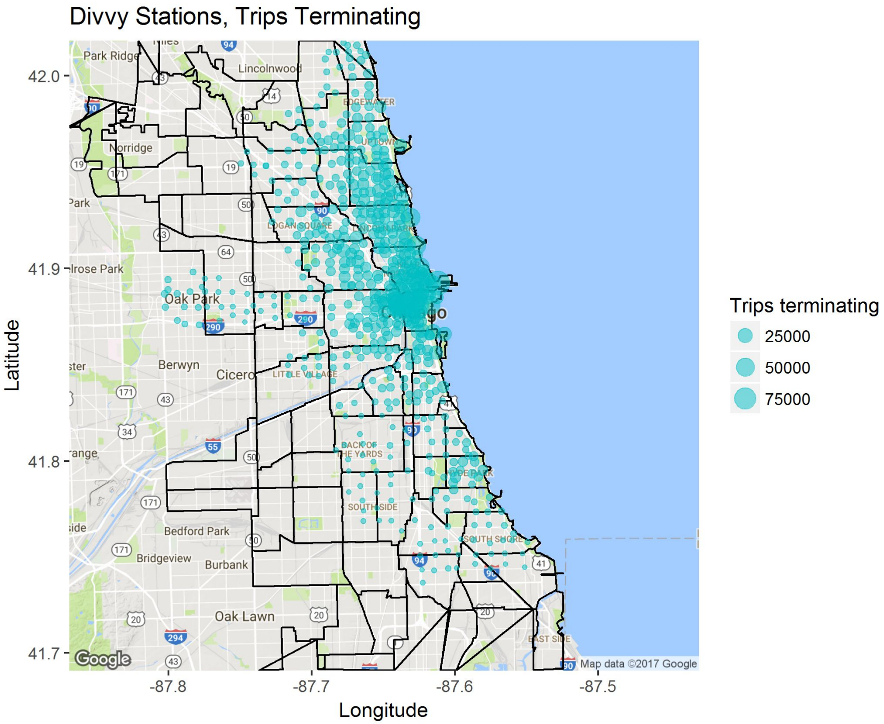
Table 2: Descriptive summary of Divvy stations

Station attributes	Mean	Standard deviation
Number of trips originating	6188.27	8487.86
Number of trips terminating	6188.27	8655.73
Station capacity	17.19	5.56
Presence of CTA stops in proximity (50m)	48.4%	
Presence of CTA stops in proximity (100m)	73.1%	
Presence of CTA stops in proximity (200m)	88.8%	
Presence of CTA stops in proximity (300m)	92.6%	
Number of CTA stops in proximity (50m)	0.91	1.19
(only stations in proximity of CTA stops)	1.88	1.05
Number of CTA stops in proximity (100m)	1.98	1.82
(only stations in proximity of CTA stops)	2.70	1.60
Number of CTA stops in proximity (200m)	4.30	2.94
(only stations in proximity of CTA stops)	4.85	2.67
Number of CTA stops in proximity (300m)	7.66	4.60
(only stations in proximity of CTA stops)	8.27	4.22



Data: Divvy stations





Model

Outcome/dependent variable Y

- Potential multi-modality of each trip for the given proximity standard
    - If potentially multi-modal, the trip starts from/ends to a station in proximity with any CTA stops about the time when bus or rail arrive to/depart from such stops
    - Within the window of 1 to 5 minutes (assuming the trip was planned to be multi-modal)
- Proximity
    - Measured in Manhattan distance
    - Comparison of multiple proximity standards
        - 50m
        - 100m
        - 200m
        - 300m (used in most previous studies; e.g. Faghih-Imani & Eluru (2015))

Model

Attributes/independent variables **x**

- Trip-level attributes (user type, gender, age, trip duration)
- Temporal attributes (day of week, time of day, temperature, precipitation)
- Geographic attributes (CBD, population density, job density)

Model

Logistic regression (for each proximity standard)

- Outcome y for trip *i*: $Pr(Y_i = y_i | p_i) = p_i^{y_i}(1 - p_i)^{(1-y_i)}$
- Probability function: $p_i = \dfrac{\exp(\beta_0 + \beta'\mathbf{x}_i)}{1 + \exp(\beta_0 + \beta'\mathbf{x}_i)}$
    - p_i Probability that the trip *i* is potentially multi-modal
    - β The vector of coefficients
    - $\mathbf{x}_i$ The vector of independent variables
- Likelihood function: $L = \prod_{i=1}^{N} p_i^{y_i}(1 - p_i)^{(1-y_i)}$

Summary

- Does Divvy offer a solution to the last mile problem for CTA rides?
- Comparison of various proximity standards

Reference

- DeMaio, P. (2009). Bike-Sharing: History, Impacts, Models of Provision, and Future. Journal of Public Transportation, 12(4), 41-56.
- Faghih-Imani, A., & Eluru, N. (2015). Analysing bicycle-sharing system user destination choice preferences: Chicago's Divvy system. Journal of Transport Geography, 44, 53-64.
- Martin, E. W., & Shaheen, S. A. (2014). Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. Journal of Transport Geography, 41, 315–324.
- Zellner, M., Massey, D., Shiftan, Y., Levine, J., & Arquero, M. J. (2016). Overcoming the Last-Mile Problem with Transportation and Land-Use Improvements: An Agent-Based Approach. International Journal of Transportation, 4(1), 1-26.

Reid McIlroy-Young

April 5, 2017

THE UNIVERSITY OF CHICAGO

Question

How do the tools used by scientists shape their work?
Specifically:

What causes new scientific software to be adopted by a community and what communities are most likely to integrate a new tool?

Data

- Scientific meta-data collections
  - Web of Science, pubmed, etc
- Open source portals
  - CRAN, Github, PyPi, etc

Method

- Identify communities / disciplines to study
  - e.g. statistics community
- Develop means of identifying new software tools
  - some subset of journals publish new software: *journal of statistical software*, *The R Journal*, *Journal of Multiscale Modelling and Simulation*
  - Pubmed has some of the linking already done
  - Generalize to other sources by hand and ML techniques

Analysis

- Can we identifying accurately publications introducing new software packages, libraries, code snippets, etc?
- How are new tools distributed within the network?
- What predicts their success, both in the literature and in usage?
- What causes them to adopted by their scientific community, by other communities and/or non-scientists?
- What communities are most frequently adopting new tools?

OSCARS:
AESTHETIC SUCCESS OR COMMERCIAL SUCCESS?

MACS 30200

YIQING ZHU

THE 67TH ACADEMY AWARDS | 1995

BEST PICTURE

WINNER

FORREST GUMP

NOMINEES

FOUR WEDDINGS AND A FUNERAL

PULP FICTION

QUIZ SHOW

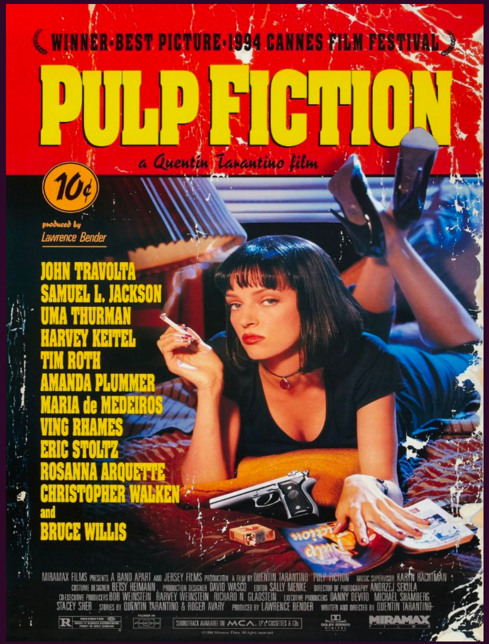
THE SHAWSHANK REDEMPTION

THE 67TH ACADEMY AWARDS | 1995

BEST PICTURE

FORREST GUMP

PULP FICTION



FORREST GUMP		PULP FICTION	
VIEWER RATING	CRITIC RATING	VIEWER RATING	CRITIC RATING
4.1 / 5	7.2 / 10	4.2 / 5	9.1 / 10
Rotten Tomatoes AUDIENCE SCORE	Rotten Tomatoes TOMATOMETER	Rotten Tomatoes AUDIENCE SCORE	Rotten Tomatoes TOMATOMETER
8.8 / 10	82	8.9 / 10	94
IMDb USER RATINGS	Metacritic METASCORE	IMDb USER RATINGS	Metacritic METASCORE

WHY DID *FORREST GUMP* WIN OSCAR?

The National Academy Claims:

- It can be regarded "Three quarters of a century of recognizing excellence in filmmaking achievement."
- Its judges consist of "the most gifted and skilled artists and craftsmen in the motion picture world."
- It warns its judges that they "may be importuned by advertisements and other lobbying tactics" and emphasizes "that excellence in filmmaking is the ONLY factor [to] consider in casting [their] votes."

WE ALL KNOW THAT IS NOT TRUE.

FORREST GUMP

BUDGET

55, 000, 000

GROSS

329, 691, 196

PULP FICTION

BUDGET

8, 000, 000

GROSS

107, 930, 000

DOES THIS MEAN THAT THE COMMECIAL SUCCESS IS MORE IMPORTANT THAN AESTHETIC SUCCESS?

RESEARCH QUESTION
How to predict The Best Picture Oscar Award Winner?

- **Construct the model of The Best Picture Oscar Award Winner for each year mainly by two categories: ART and BUSINESS**
- **Capture the annual variation of the model used to predict The Best Picture Oscar Award Winner for each year**
- **Aggregate and summarize past models so as to predict The Best Picture Oscar Award Winner in the following year**

BEST PICTURE OSCAR PREDICTIVE MODEL

ART

- **Critic Review**
- **Critic Rating**

BUSINESS

- **Viewer Review**
- **Viewer Rating**
- **Budget / Cast Size**
- **Gross**
- **Relationship to Hot Topics**

The process of movie making is both an industry and an art.

-Hammad Afzal (2016)

Philosophers typically put the burden of proving quality on experts, while economists often argue that the actual choices made by consumers are a better measure.

-Landes (2002)

DATASET

All U.S. movies available on IMDb and Rotten Tomatoes since 1929

Analyses & Computational tools

- **Python**
- **SQL**
- **Content analysis**
  - NLTR, etc.
- **Machine learning algorithms:**
  - SVM
  - Logistic Regression
  - Random Forest, etc.

POTENTIAL CONTRIBUTION

- More accurately predict Best Picture Oscar Award Winner
- Observe the development of Hollywood movie industry and forecast the trend

A Perceptual Map of the Decision Making

Perspectives on Computational Research | Apr 5, 2017 | HyungJin Cho

INDEX

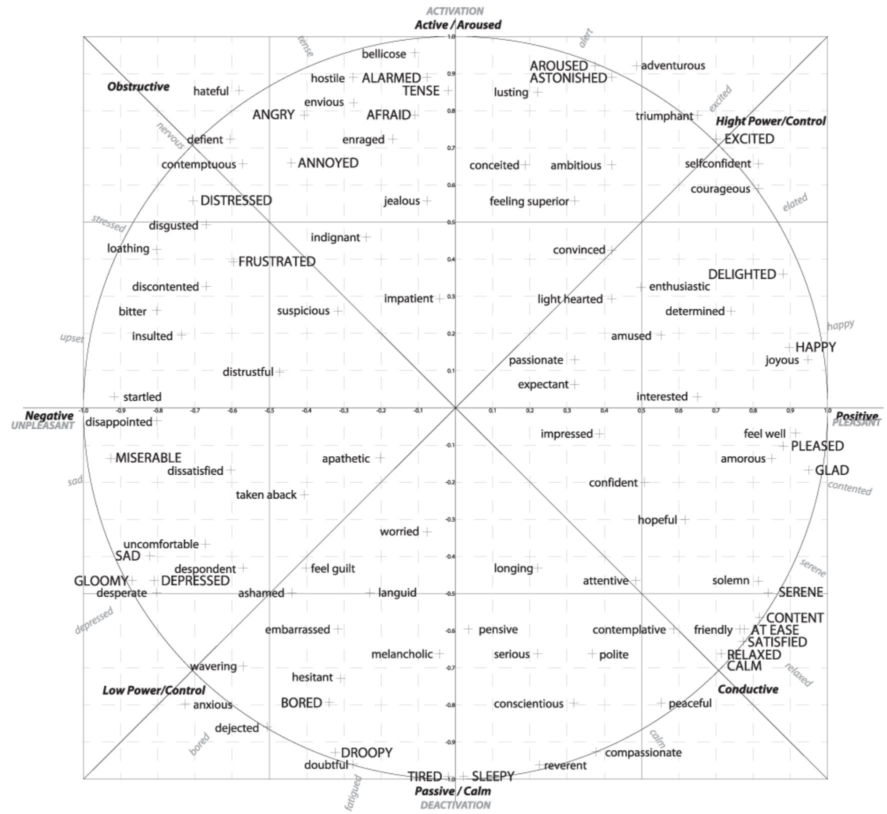
1. Reference
2. Conclusion
3. Method
4. Abstract

Reference

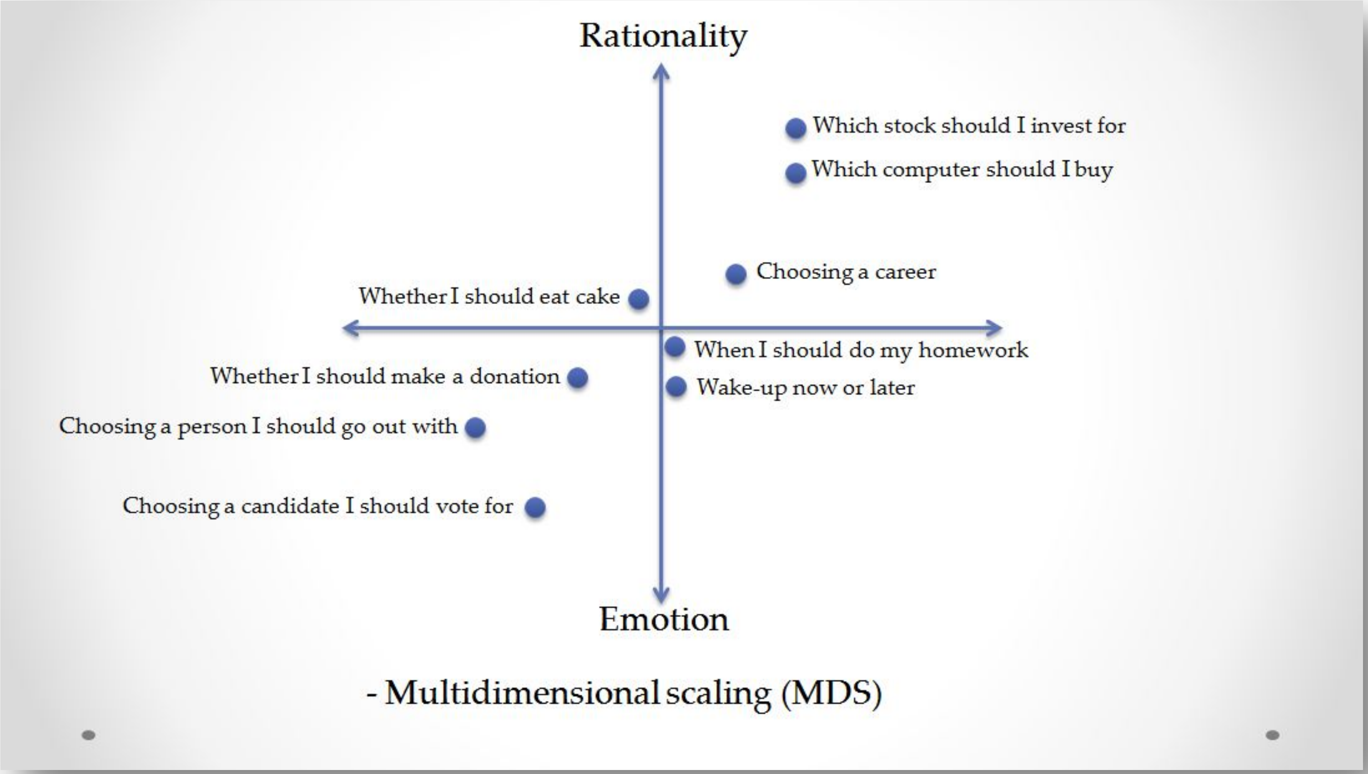
Campbell, N. R. (1920). Foundations of Science (formerly titled: Physics, The Elements).
Russell, J. A., & Pratt, G. (1980). A description of the affective quality attributed to environments. Journal of personality and social psychology, 38(2), 311.
Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. Psychological bulletin, 98(2), 219.
Giguère, G. (2006). Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using SPSS. Tutorials in Quantitative Methods for Psychology, 2(1), 27-38.
Mano, H., & Oliver, R. L. (1993). Assessing the dimensionality and structure of the consumption experience: evaluation, feeling, and satisfaction. Journal of Consumer research, 20(3), 451-466.
Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. Cognition & Emotion, 14(4), 473-493.
Barrett, L. F. (2006). Are emotions natural kinds?. Perspectives on psychological science, 1(1), 28-58.
Trewatha, R. L., & Newport, M. G. (1979). Management, functions and behavior (3rd ed.). Dallas, TX: Business Publications, Inc.

Russell, J. A., & Pratt, G. (1980)

Previous research suggests that a concept whose defining attribute defies description can be studied using multidimensional scaling.



Conclusion



> **The underlying dimensions extracted from the spatial configuration of the data are thought to reflect the hidden structures, or important relationships, within it.**
> **(Young & Hamer, 1987)**

Method

Regression Model & Multidimensional Scaling

- Multidimensional Scaling is a mean of visualizing the level of similarity of individual cases of a dataset.
- Pairwise similarities reconstruct a map that preserves distances.

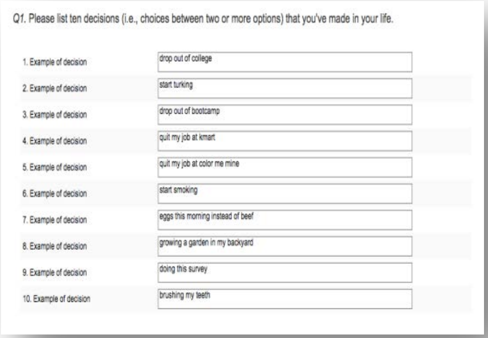
$$\Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,I} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,I} \\ \vdots & \vdots & & \vdots \\ \delta_{I,1} & \delta_{I,2} & \cdots & \delta_{I,I} \end{pmatrix}.$$

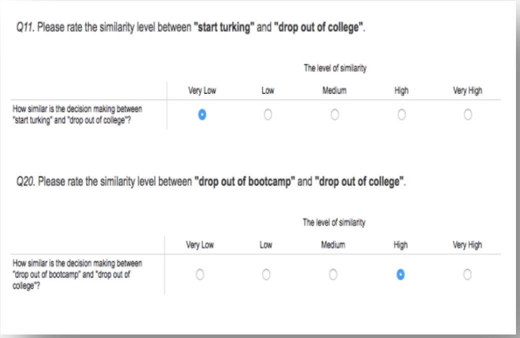
Data & Participants

- 120 people are to be recruited through Amazon Mturk to participate in a survey.
- Demographic questionnaire is to be included in a survey.

Method

Procedure





5.00	.	.	.	.	.	.	.	.	.
1.00	5.00	.	.	.	.	.	.	.	.
4.00	1.00	5.00	.	.	.	.	.	.	.
3.00	3.00	3.00	5.00	.	.	.	.	.	.
4.00	3.00	3.00	5.00	5.00	.	.	.	.	.
2.00	2.00	1.00	1.00	2.00	5.00	.	.	.	.
1.00	1.00	1.00	1.00	1.00	1.00	5.00	.	.	.
1.00	1.00	1.00	1.00	1.00	1.00	2.00	5.00	.	.
1.00	4.00	1.00	1.00	1.00	1.00	2.00	1.00	5.00	.
1.00	1.00	1.00	1.00	1.00	1.00	3.00	1.00	2.00	5.00

Step1. List of decisions
In this stage, participants list ten decisions that they've made in their life.

Step2. Similarity between decisions
Participants are asked how one decision is similar to another.

Step3. Computational Analysis
MDS analysis is conducted using function cmdscale() in R.

Abstract

A Perceptual Map of the Decision Making

What is the underlying dimensions in diverse types of decisions in the level of people’s perception? In this study, MDS is used to create a map of decision-making and to extract the structure and relationships within it.

Keyword: Decision Making, Multidimensional Scaling (MDS)

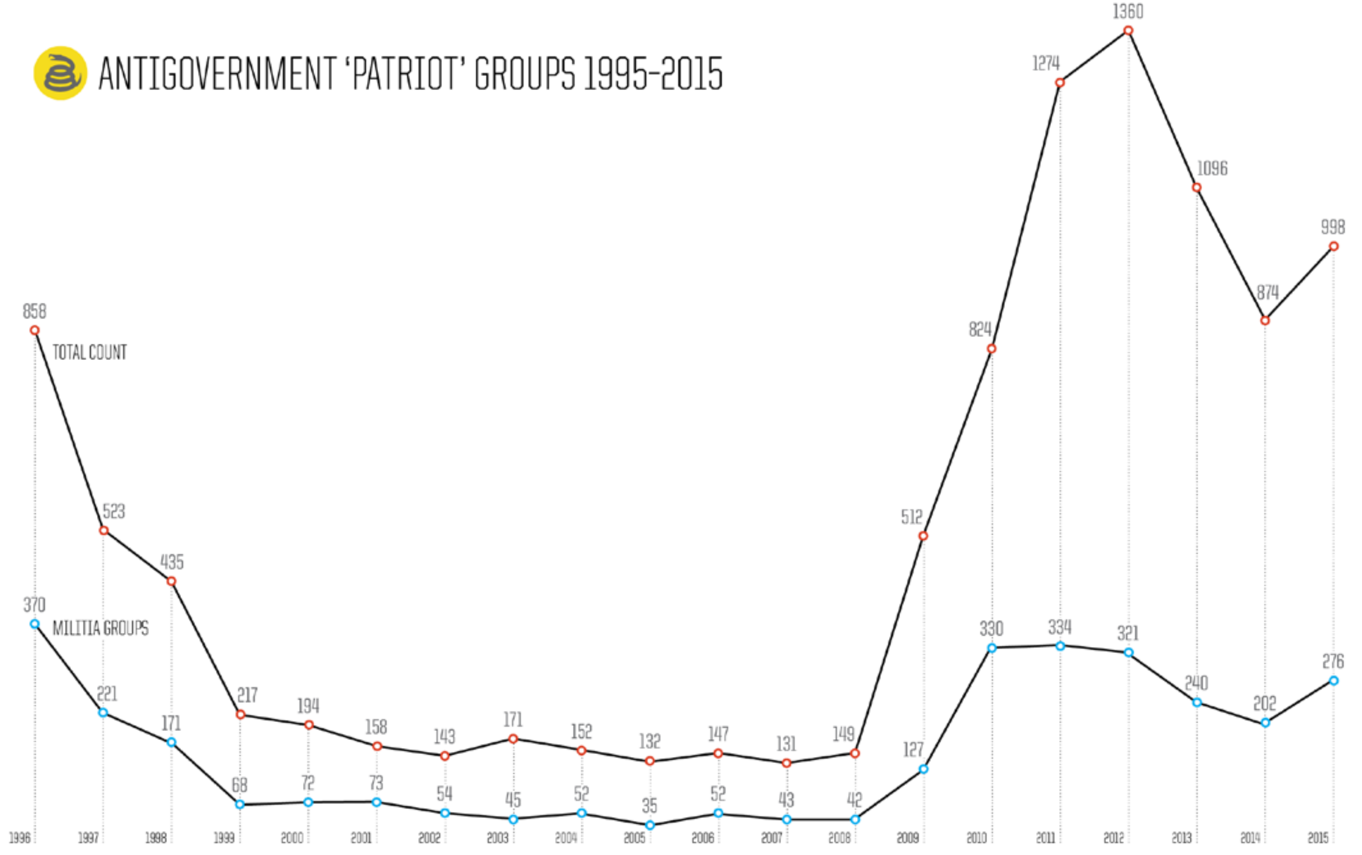
SIGNIFICANT FACTORS IN ANTIGOVERNMENT GROUP SOCIAL MOBILIZATION

AN APPLICATION OF QUANTITATIVE METHODS FOR HISTORICAL ANALYSIS

JOANNA TUNG

WHAT HISTORICAL AND DEMOGRAPHIC TRENDS APPEAR PREDICTIVE OF ANTIGOVERNMENT GROUP MOBILIZATION BY COUNTY DURING THE PERIOD FROM 1992-2014?

- Existing Theories
  - Material conditions
    - Employment, Income
  - Structural-Cultural conditions
    - Race, Gender, Degree of Urbanization
  - Psychological-Spiritual conditions
    - Religiosity, Veteran Status
  - Disruption
    - Influx of immigrants, war, political turmoil



Credit: Southern Poverty Law Center, https://www.splcenter.org/active-antigovernment-groups-united-states

DATA SOURCES

- U.S. Census Bureau – American Community Survey
  - Veteran population by war
  - Foreign-born population
  - Employment
  - Education
  - Race
- Southern Poverty Law Center (1992-2015)
  - Antigovernment group by city
  - Militia (yes/no)
- Interuniversity Consortium for Political and Social Research (ICPSR)
  - Uniform Crime Reporting (FBI Hate Crime Statistics)
  - FIPS County Code
- Association of Religious Data Archives
  - Judeo-Christian church number and membership

BUILDING THE NARRATIVE:
METHODS AND APPROACHES

- Introducing historical time
  - Add variable that accounts for a change in time
  - "Rolling" Assessment

- Classification Problem – County Level
  - No Antigovernment Group
  - Antigovernment Group (non-militia)
  - Antigovernment Group (militia)

- Exploratory: Principle Component Analysis (PCA)
- Significant variables for producing accurate models
  - Decision Trees
  - Logistic Regression

The Effect of Submission and Comment History on Opinion Malleability in /r/changemyview.

- Changing a person's opinion is a common task in a variety of settings.
- /r/changemyview offers a forum where the reasoning behind opinions must be stated and successful arguments must be explicitly rewarded.
- Tan et al. attempted to determine the malleability of an opinion from the way the original poster (OP) presented his or her reasoning.
  - However, in many cases the OP also has previous Reddit participation through comments and submissions.
  - The feature set in Tan et al. can be expanded to past the initial submission.

Deltas awarded in "CMV: FIRE BAD!" (self.DeltaLog)

submitted 1 day ago * (last edited 14 hours ago) by DeltaBot [M]

Below is a list of the deltas awarded in this post[1].

Please note that a change of view is not necessarily a reversal[2], and that OP awarding a delta doesn't mean the conversation has ended.

For a full explanation of the delta system, see here[3].

Deltas from OP /u/theshantanu[4]

- 1 delta from OP to /u/XXX69694206969XXX[5] for "[Quote] USE FIRE IN CAVE NO NEED SKIN IN CAVE COMFORTABLE AND WARM DIP SKIN IN ANIMAL FAT WRAP SKIN ROUND STICK USE FIRE ON SKIN SKIN BURN BUT NO STICK BURN FOR LONG TIME [Quote] ONLY LIKE RAW BECAUSE...[6]"
- 1 delta from OP to /u/ubbergoat[7] for "ANIMAL NO CONTROL FIRE. THEY NOT SMART. YOU MAN, MAN CONTROL FIRE. MAKES FOR SMART MAKING. YOU MUCH BETTER THEN ANIMAL WITH FIRE.[8]"
- 1 delta from OP to /u/StockingSaboteur[9] for "SNOW COLD. LONG NIGHTS COLDER. EVEN WITH SKINS BABBY DIE. WITH FIRE BABBY NO DIE. FIRE GOOD. FIRE MAKE MEAT GOOD. WITH NO FIRE MEAT HURT BELLY. WITH FIRE NO HURT. EAT MORE MEAT. BE BIGGER THAN NO FIR...[10]"
- 1 delta from OP to /u/sirgippy[11] for "[Quote] WOLVES TOUGH TO TRAIN. SOMETIMES WOLVES EAT BABIES. MUST FEED WOLVES MEAT LIKE FEED MAN (OR ELSE WOLVES EAT MORE BABIES). NOT MUCH CONTROL OF WOLF JUST LIKE NOT MUCH CONTROL OF FIRE. FIRE DOES...[12]"

Data and Methods

- Tan et al. Focused on 12,272 cases where at least 10 challengers attempted counterarguments, and where the OP replied at least once.
  - Data ranged from 1/1/13-9/1/15
- I will experiment with similar restrictions, but will utilize data from 1/1/13-12/31/16.
  - I will utilize logistic regression and/or other binary classification methods where I can extract variable importance measures.
  - Does self-affirmation theory extend to past Reddit Participation?
  - Does a notion of reciprocation cause Ops to give Δ's?

/u/theshantanu has received 10 deltas:

Date	Submission	Delta Comment	Awarded By
3/24/2017	CMV: It's a terrible idea to live in Alaska	Link	/u/Dat_Das
2/9/2017	CMV: Flooding is not a natural disaster, but human stupidity.	Link	/u/Garlicplanet
12/20/2016	CMV: political parties bring nothing but problems to the political system, and only serve to divide people and make government slower.	Link	/u/QueenCharla
12/18/2016	CMV: The single best improvement to be made to /r/aww is to remove the titles.	Link	/u/MSPaintClock
7/18/2016	CMV: There should be a legal requirement for public clocks to be on time.	Link	/u/NirvanaFighter
5/27/2016	CMV: People that use tragic events in order to convey their ideas are selfish.	Link	/u/Vitrin99
5/25/2016	CMV:If I am part of a conversation, I should be allowed to record it without informing the others	Link	/u/Impacatus
3/21/2016	CMV: Wall Street are literally terrorists	Link	/u/bobdylan401
2/25/2016	CMV: The depressing okcupid/dating studies tells black women we should use skin bleaching creams to lighten our skin if we want to find a date.	Link	/u/JubbyO
2/24/2016	CMV: As a short male, growing taller through limb lengthening surgery would be beneficial for me.	Link	/u/YoungandEccentric

MA CSS Research Topic

Benjamin Rothschild

Research Question

How accurate can I create a model to predict locations of homicides in Chicago?

Background

Predictive Policing - the concentration of police resources in stable crime hotspots has proven effective in reducing crime, but the extent to which police can monitor changing crime hotspots is unknown.

In Los Angeles a model was *1.4 - 2.2* times better at predicting crime compared to a dedicated crime analyst and with this improved performance *7.4%* of crime was reduced in a controlled experiment that used a model to predict hotspots.

In this study the prediction algorithm used historical crime data to predict future crime hotspots (did not rely on census or demographic data)

Randomized Controlled Field Trails for Predictive Policing

G. O. Mohler, M.B. Short, Sean Malinowski, Mark Johnson, et al.

Background

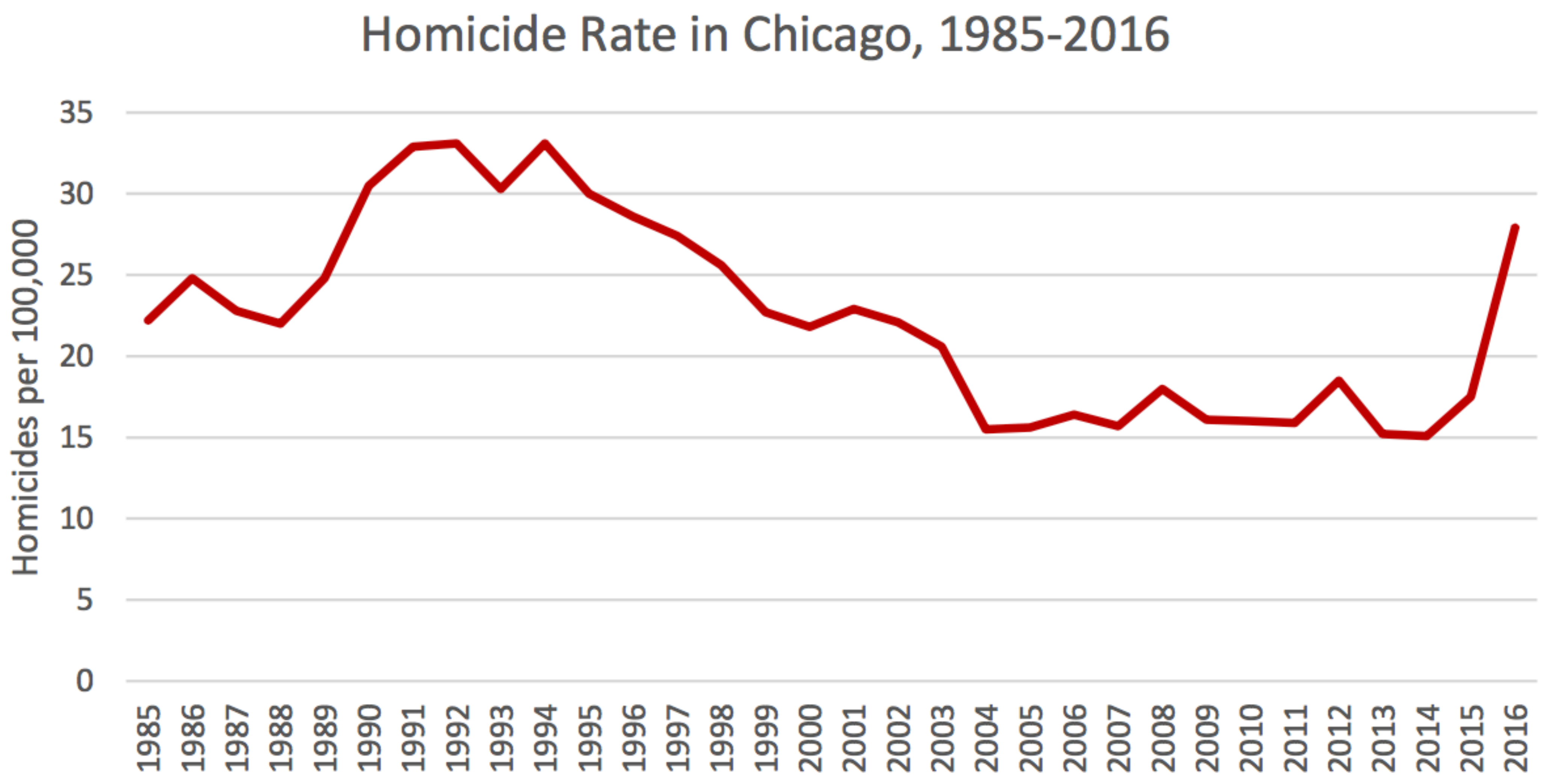
Recent social science research has taken advantage of "Big Data" sources to achieve more accurate predictions and measurements of real-life phenomena.

For example Google Street View images were successfully used to measure income in New York City and Boston.

Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life

Edward L. Glaeser, Scott Duke Kominers, Michael Luca, Mikhil Naik

Datasets



Homicide Rate in Chicago, 1985-2016
Homicides per 100,000
35
30
25
20
15
10
5
0
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016

Datasets

- Census Tracts
  - 5 Year estimates
  - Data on Age, Income, Housing, etc
- Community Input Data (311 Calls)
  - Updated Daily
  - Service Requests like abandoned buildings, disturbances
  - Location, time, request type
- Chicago Crime Data
  - Monthly Data
  - Crime type, location, time
  - Lower-level crime data
- Zillow
  - Rental Price Data
- Twitter
  - Real-time or can query tweets by place

Methods

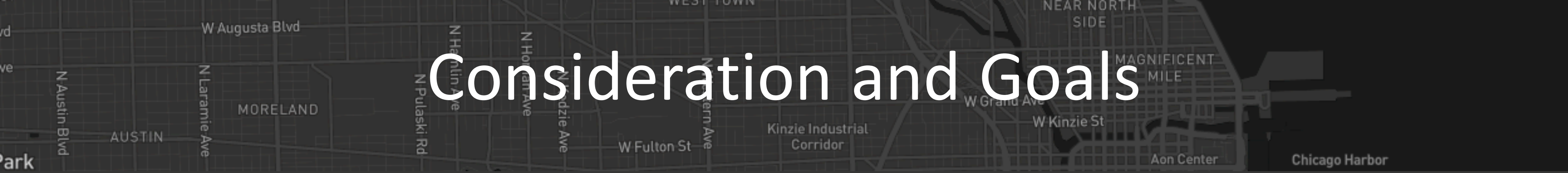
- Neural Network
- Decision Trees
- Support Vector Machines
- Logit/Probit Regression

Evaluation Methods

- I expect the output of my model to give a score for specific areas (blocks/radius/census tracts, etc) and I will try to correlate these scores with the data from the Chicago Crime Dataset.

- Evaluation Criteria:
  - Error Rate
  - Proportional Reduction in Error (PRE)
  - Area Under the Curve (AUC)

Consideration and Goals



- I want to try to make my geographic units of analysis as small as possible
  - How does my error rate changes with different boundary sizes

- Similarly I want to make my unit of time analysis as small as possible
  - How often can I make a prediction? How does this affect the error rate?

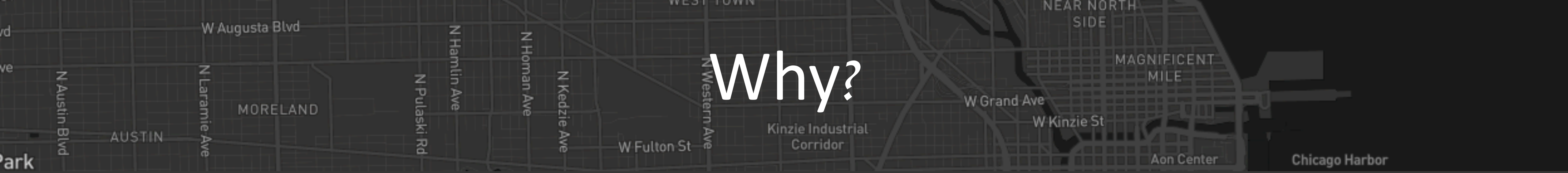
- What datasets are most informative in helping my prediction?
  - How does adding datasets affect the Error rate?

Possible Conclusions

- I can predict there will be a crime on a .5 mile radius and 5 hour timespan with at 15% success rate

- I can predict there will be a crime on a 2 block radius and 10 hour timespan at 30% success rate

- I can predict nothing ☹



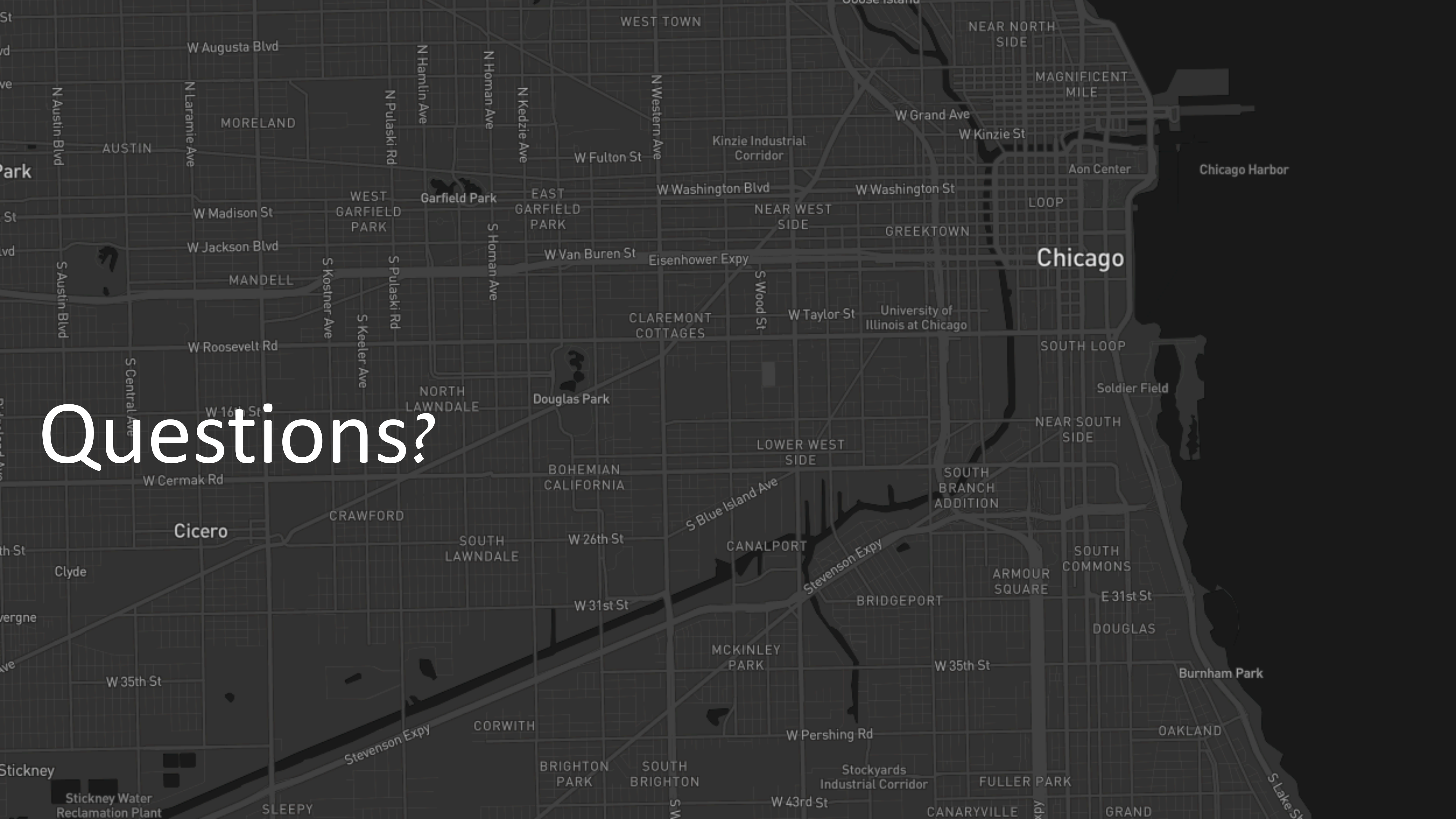
Why?

- A model like this could help with resource allocation with police department or with community investment initiatives
- Analysis like this can lead to further questions about what environments lead to crime
- By seeing how the model improves or not when data sources are added I can hypothesize about what other datasets would be helpful in improving my model

References and Resources?

- Datasets
  - https://censusreporter.org/profiles/86000US60657-60657/
  - https://data.cityofchicago.org/Service-Requests/311-Service-Requests-Vacant-and-Abandoned-Building/7nii-7srd/data
  - https://data.cityofchicago.org/view/5cd6-ry5g
  - https://www.zillow.com/research/zillow-rent-index-methodology-2393/
  - https://dev.twitter.com/rest/public/search-by-place
- Articles
  - Glaeser, Edward L., Scott Duke Kominers, Michael Luca, and Nikhil Naik. "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life." (pdf)Economic Inquiry (forthcoming). View Details
- Websites
  - https://www.civicscape.com

Questions?



WEST TOWN
NEAR NORTH SIDE
MAGNIFICENT MILE
W Augusta Blvd
N Austin Blvd
N Laramie Ave
MORELAND
AUSTIN
N Pulaski Rd
N Hamlin Ave
N Homan Ave
N Kedzie Ave
N Western Ave
W Fulton St
Kinzie Industrial Corridor
W Grand Ave
W Kinzie St
Aon Center
Chicago Harbor
W Washington Blvd
W Washington St
WEST GARFIELD PARK
Garfield Park
EAST GARFIELD PARK
W Madison St
NEAR WEST SIDE
LOOP
GREEKTOWN
W Jackson Blvd
W Van Buren St
Eisenhower Expy
Chicago
MANDELL
S Austin Blvd
S Kostner Ave
S Pulaski Rd
S Homan Ave
S Wood St
W Taylor St
University of Illinois at Chicago
CLAREMONT COTTAGES
S Keeler Ave
W Roosevelt Rd
SOUTH LOOP
S Central Ave
Soldier Field
NORTH LAWNDALE
Douglas Park
W 16th St
NEAR SOUTH SIDE
LOWER WEST SIDE
W Cermak Rd
BOHEMIAN CALIFORNIA
SOUTH BRANCH ADDITION
S Blue Island Ave
CRAWFORD
Cicero
SOUTH LAWNDALE
W 26th St
CANALPORT
Stevenson Expy
SOUTH COMMONS
ARMOUR SQUARE
Clyde
BRIDGEPORT
E 31st St
W 31st St
DOUGLAS
MCKINLEY PARK
W 35th St
Burnham Park
W 35th St
CORWITH
OAKLAND
Stevenson Expy
W Pershing Rd
Stickney
BRIGHTON PARK
SOUTH BRIGHTON
Stockyards Industrial Corridor
FULLER PARK
Stickney Water Reclamation Plant
SLEEPY
W 43rd St
CANARYVILLE
GRAND

Research Proposal

Yuqing (Candice) Zhang

Background

- Few studies have evaluated why restaurants failed
- Location, affiliation, and size are significant influences on restaurants' mortality[2]
  - But, to what extent?
  - Cofounding variables?

Research Question

- How do the mechanisms of food, ambience and location impact the **closing** of restaurants, which got their license in 2012 in Chicago area?

Background-Yelp Reviews

- Yelp is currently the most popular online consumer review website used for local business reviews and recommendations
- By the end of Q2 2016, yelpers have written more than 108 million reviews.
- 82 percent say their purchase decisions have been directly influenced by online reviews
- A one star increase in Yelp rating leads to a 5-9 percent increase in revenue[1]

Data

- **YELP**
  - Business
  - Reviews
- **Chicago Business License**

Business Data

yelp_academic_dataset_business.json

```
{
    "business_id":"encrypted business id",
    "name":"business name",
    "neighborhood":"hood name",
    "address":"full address",
    "city":"city",
    "state":"state -- if applicable --",
    "postal code":"postal code",
    "latitude":latitude,
    "longitude":longitude,
    "stars":star rating, rounded to half-stars,
    "review_count":number of reviews,
    "is_open":0/1 (closed/open),
    "attributes":["an array of strings: each array element is an attribute"],
    "categories":["an array of strings of business categories"],
    "hours":["an array of strings of business hours"],
    "type": "business"
}
```

Reviews

yelp_academic_dataset_review.json

```
{
    "review_id":"encrypted review id",
    "user_id":"encrypted user id",
    "business_id":"encrypted business id",
    "stars":star rating, rounded to half-stars,
    "date":"date formatted like 2009-12-19",
    "text":"review text",
    "useful":number of useful votes received,
    "funny":number of funny votes received,
    "cool": number of cool review votes received,
    "type": "review"
}
```

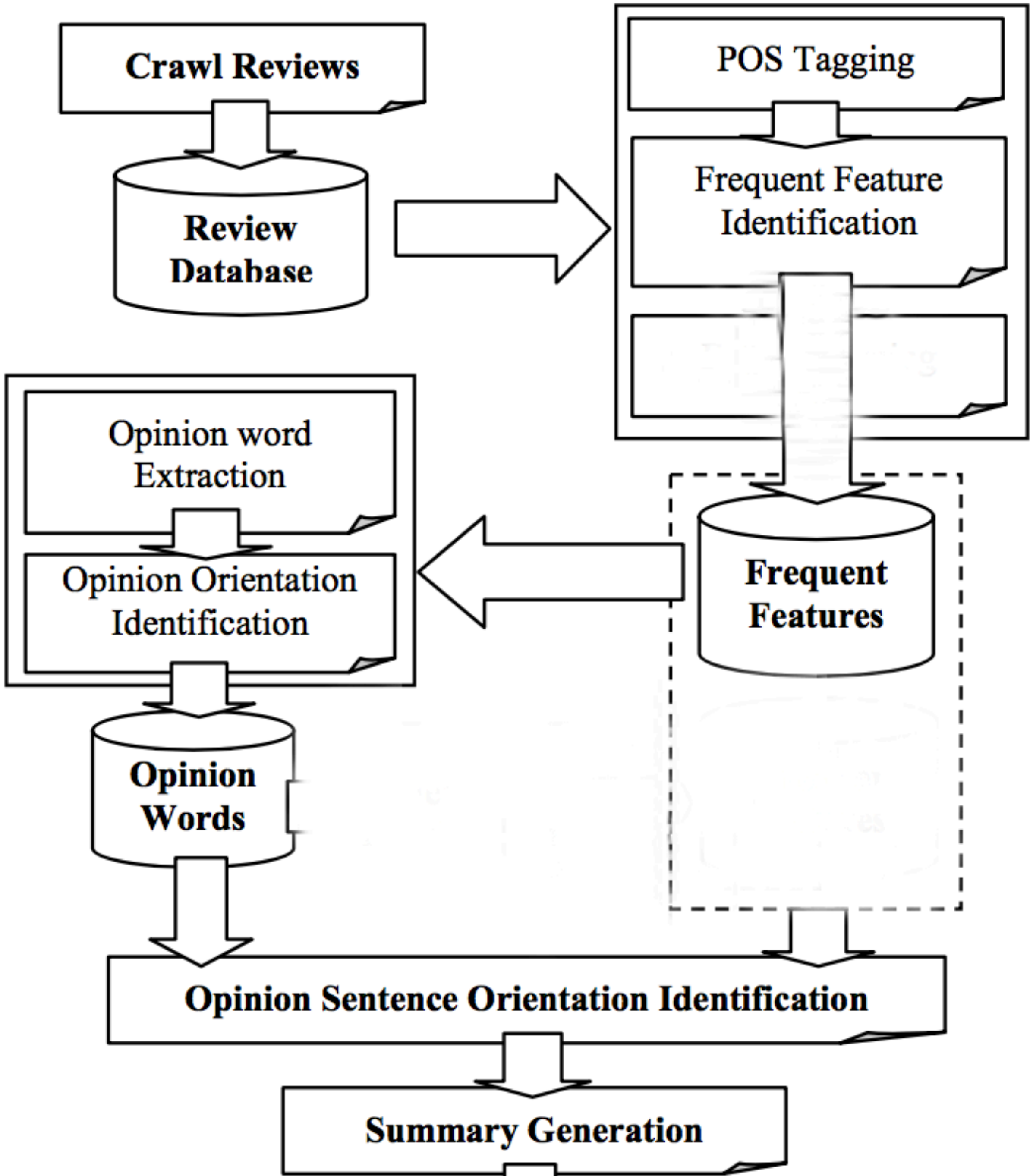
License Issue Date

LEGAL NAME
BELL OIL TERMINAL INC
BUCCI BIG & TALL INC.
PROJECT: VISION , INC.
FOLASHADE'S CLEANING SERVICE INC.
WALGREEN CO.
BURKS HEATING AND COOLING SOLUTIONS, LLC
BELL OIL TERMINAL INC
JAM PRODUCTIONS, LTD.
ANGELINE R. MC CARTHY
REVOLUTION BREWING, LLC
BELL OIL TERMINAL INC
WALGREEN CO.

DATE ISSUED
8/11/06
8/30/16
6/22/16
4/1/16
5/11/07
8/30/16
4/16/04
8/30/16
8/30/16
8/30/16
3/5/04
4/28/03
8/30/16
8/30/16
8/30/16

Methodology - Filter

- Include only
  - Restaurants
  - Chicago area
  - Got their license issued at 2012
  - Is closed
  - Split reviews into year 2012,2013,2014,2015,2016,2017



The Proposed Techniques

Methodology-Frequent Feature Identification

- Category Prediction
  - Trains on review data and generates a simple naïve-Bayes model that can predict the category of some text

`category_predictor`: Given some text, predict likely categories. For example:

```
$ python category_predictor/category_predictor.py yelp_academic_dataset.json > category_predictor.json
$ python category_predictor/predict.py category_predictor.json "bacon donut"
Category: "Food" - 82.66% chance
Category: "Restaurants" - 16.99% chance
Category: "Donuts" - 0.12% chance
Category: "Basque" - 0.02% chance
Category: "Spanish" - 0.02% chance
```

Methodology-Sentiment Analysis

- Natural Language Tool Kit

- **POS Tagging**:
  - NLProcessor linguistic parser
  - Split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc)

- **FFI** :
  - Food, location, ambience
  - Category Prediction Function from Yelp

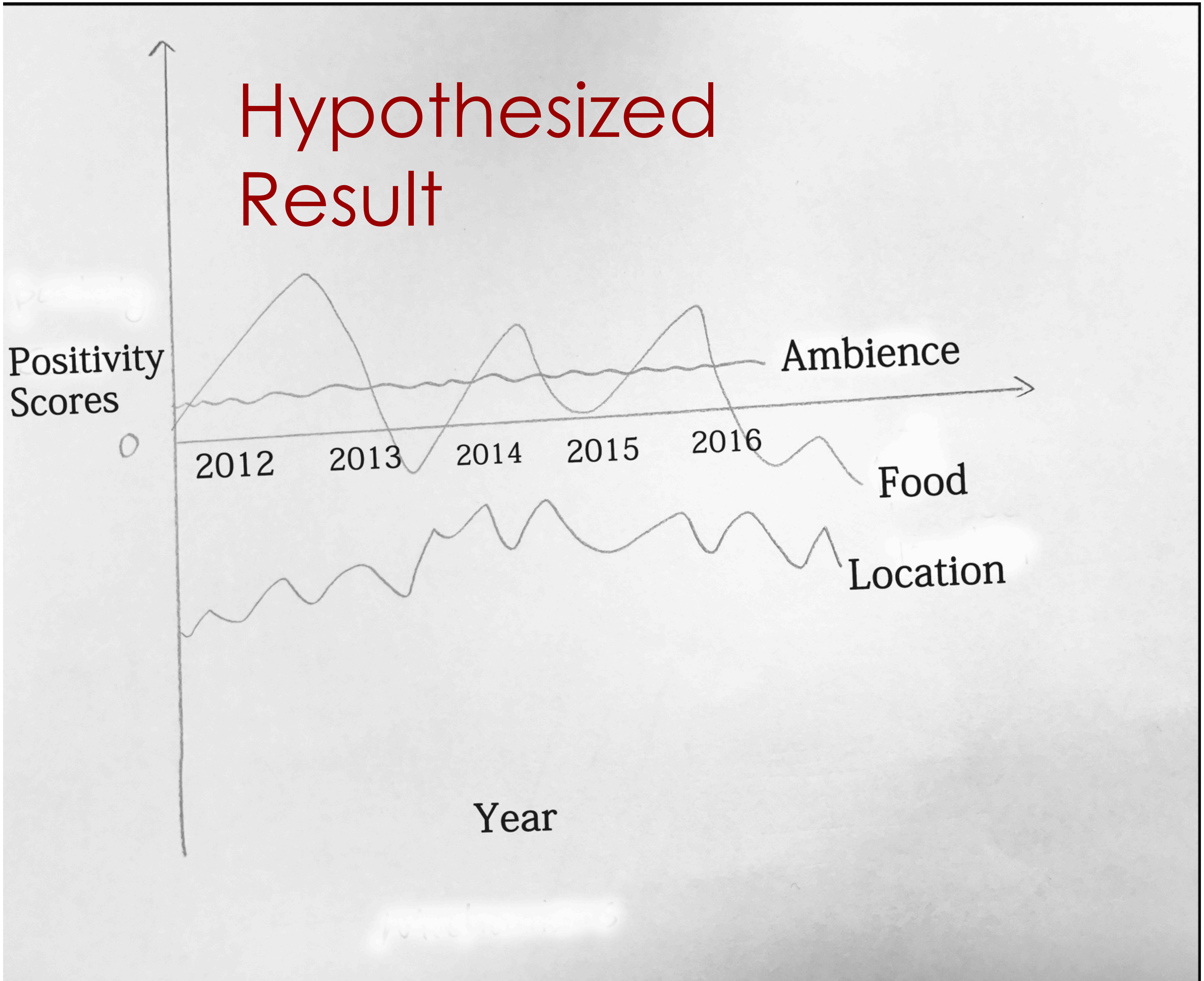
- **Opinion Words Extraction**:
  - Adjectives is useful for predicting whether a sentence is subjective, i.e., expressing an opinion.
  - Limit the opinion words extraction to those sentences that contain one or more product features

Methodology-Sentiment Analysis

- **Orientation Identification for Opinion Words**:
  - Sentiment Intensity Analyzer function from NLTK to calculate a positivity score for each word

- **Predicting the Orientations of Opinion Sentences**:
  - Use the dominant orientation of the opinion words in the sentence to determine the orientation of the sentence

Hypothesized Result



Reference

- [1]Luca, Michael. "Reviews, Reputation, and Revenue: The Case of Yelp.com.' Harvard Business School Working Paper, No. 12-016, September 2011. (Revised March 2016. Revise and resubmit at the *American Economic Journal – Applied Economics*.)

- [2]Parsa, H. G., Self, J., Sydnor-Busso, S., & Yoon, H. J. (2011). Why Restaurants Fail? Part II - The Impact of Affiliation, Location, and Size on Restaurant Failures: Results from a Survival Analysis. Journal of Foodservice Business Research, 14(4), 360-379. doi: 10.1080/15378020.2011.625824

Residential School Attendance and Negative Adult Outcomes

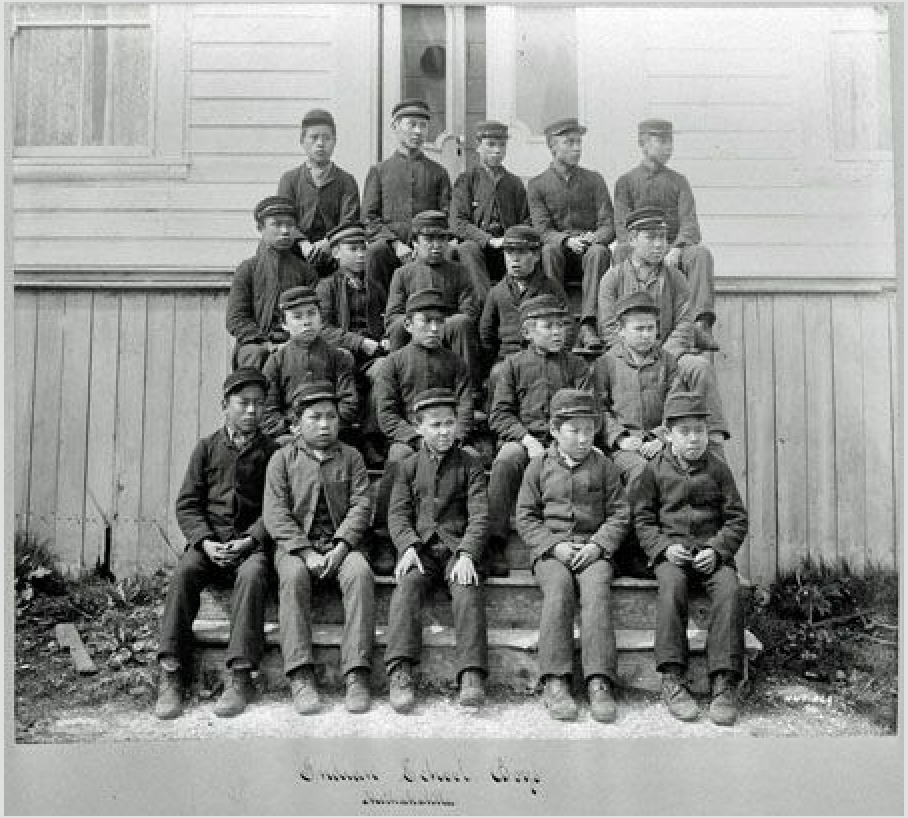
Chelsea Ernhofer

Research Question

To what extent can the psychological theory of adverse childhood experiences (ACEs) be applied to Canadian First Nations residential school attendees in order to explain negative social and psychological outcomes?

Background 🇨🇦

- History of Residential Schools
- First Nations outcomes and general statistics
- Truth and Reconciliation Commission



Past Research

ACEs

Adverse experiences during childhood are linked with several negative adult outcomes:

- ➔ Depression and mental health (Chapman et al. 2004; Edwards et al. 2003; Dahl et al. 2017)
- ➔ Learning disabilities (Felitti et al. 1998)
- ➔ Alcohol and Drug abuse (Anda et al. 2002; Dube et al. 2003)
- ➔ Early deaths and Suicide attempts (Dube et al. 2003; Afifi et al. 2008; Danese et al. 2009)

First Nations

Residential School attendance has been previously correlated with negative adult outcomes:

- ➔ Suicidal thoughts and attempts (Elias et al. 2012)
- ➔ Poor health and quality of life (Barton et al. 2005)
- ➔ Educational attainment (Barnes et al. 2006)

Current research has not yet explored the mechanisms behind these outcomes.

Data and Proposed Methods

Data

Over 700 interview transcripts taken by the Truth and Reconciliation Commission in Canada

Taken from 2008-2015

Never before used in statistical/quantitative analysis

Proposed Methods - Text Analysis & General Linear Models

Text Analysis

- ➔ Detect ACEs and key demographic variables (Pennebaker and Stone 2003; Mehl et al. 2006; Newman et al. 2008)
- ➔ Evaluate social and psychological state indicators:
  - **Emotionality** (Blonder et al. 2005; Djikic et al. 2006)
  - **Depression** (Arguello et al. 2006; Baddeley and Singer 2008)
  - **Social connections to group** (Pressman and Cohen 2007)
  - **Education level/social class** (Guastella and Dadds 2006; Centerbar et al. 2008)

GLMs

- ➔ Measure the relationship between ACEs/demographic variables and adult outcomes
  - Test whether increased exposure to adverse experiences in Residential Schools leads to poor emotional/psychological state
- ➔ Benefits of GLMs:
  - Ability to measure a potentially nonlinear relationship
  - Finds/explains relationships between variables

Questions?

References

Afifi, Tracie O., et al. "Population attributable fractions of psychiatric disorders and suicide ideation and attempts associated with adverse childhood experiences." American journal of public health 98.5 (2008): 946-952.

Anda, Robert F., et al. "Adverse childhood experiences, alcoholic parents, and later risk of alcoholism and depression." Psychiatric services 53.8 (2002): 1001-1009.

Arguello, Jaime, et al. "Talk to me: foundations for successful individual-group interactions in online communities." Proceedings of the SIGCHI conference on Human Factors in computing systems. ACM, 2006.

Baddeley, Jenna L., and Jefferson A. Singer. "Telling losses: Personality correlates and functions of bereavement narratives." Journal of Research in Personality 42.2 (2008): 421-438.

Barnes, Rosemary, Nina Josefowitz, and Ester Cole. "Residential schools: Impact on Aboriginal students' academic and cognitive development." Canadian Journal of School Psychology 21.1-2 (2006): 18-32.

Barton, Sylvia S., et al. "Health and quality of life of Aboriginal residential school survivors, Bella Coola Valley, 2001." Social Indicators Research 73.2 (2005): 295-312.

Blonder, Josip, et al. "Quantitative profiling of the detergent-resistant membrane proteome of iota-b toxin induced vero cells." Journal of proteome research 4.2 (2005): 523-531.

Centerbar, David B., et al. "Affective incoherence: when affective concepts and embodied reactions clash." Journal of personality and social psychology 94.4 (2008): 560.

Chapman, Daniel P., et al. "Adverse childhood experiences and the risk of depressive disorders in adulthood." *Journal of affective disorders* 82.2 (2004): 217-225.

Dahl, Signe Kirk, et al. "Early adversity and risk for moderate to severe unipolar depressive disorder in adolescence and adulthood: A register-based study of 978,647 individuals." *Journal of Affective Disorders* 214 (2017): 122-129.

Danese, Andrea, et al. "Adverse childhood experiences and adult risk factors for age-related disease: depression, inflammation, and clustering of metabolic risk markers." Archives of pediatrics & adolescent medicine 163.12 (2009): 1135-1143.

Djikic, Maja, Keith Oatley, and Jordan B. Peterson. "The bitter-sweet labor of emoting: The linguistic comparison of writers and physicists." Creativity research journal 18.2 (2006): 191-197.

Dube, Shanta R., et al. "The impact of adverse childhood experiences on health problems: evidence from four birth cohorts dating back to 1900." Preventive medicine 37.3 (2003): 268-277.

Dube, Shanta R., et al. "Childhood abuse, neglect, and household dysfunction and the risk of illicit drug use: the adverse childhood experiences study." Pediatrics 111.3 (2003): 564-572.

Edwards, Valerie J., et al. "Relationship between multiple forms of childhood maltreatment and adult mental health in community respondents: results from the adverse childhood experiences study." American Journal of Psychiatry 160.8 (2003): 1453-1460.

Elias, Brenda, et al. "Trauma and suicide behaviour histories among a Canadian indigenous population: an empirical exploration of the potential role of Canada's residential school system." Social science & medicine 74.10 (2012): 1560-1569.

Felitti, Vincent J., et al. "Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study." *American journal of preventive medicine* 14.4 (1998): 245-258.

Guastella, Adam J., and Mark R. Dadds. "Cognitive-behavioral models of emotional writing: A validation study." Cognitive Therapy and Research 30.3 (2006): 397-414.

Pressman, Sarah D., and Sheldon Cohen. "Use of social words in autobiographies and longevity." Psychosomatic Medicine 69.3 (2007): 262-269.

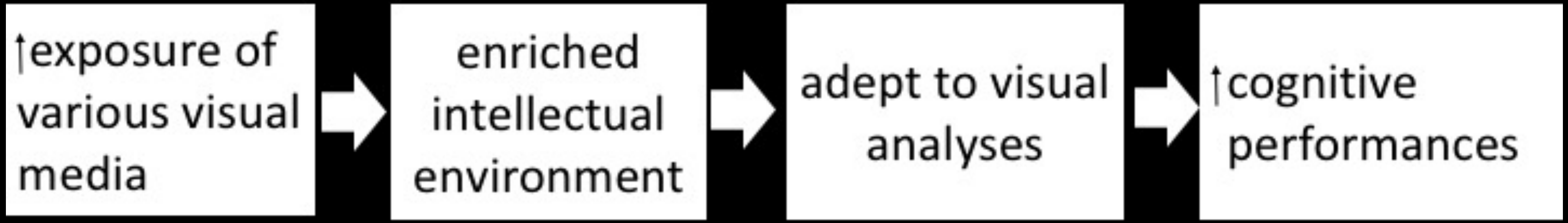
Predictive Effects of Optical Exposure on Cognitive Performances

Wenxi Xiao

April 6th 2017

Background

- Flynn effect: massive gains in IQ over time (Flynn, 1984)
- Within fluid intelligence (Rodgers & Wänströ, 2007)
- Optical exposure theory (Greenfield, 1998)



  - Conflicting results (e.g., Greenfield, 1998; Dickens & Flynn, 2002)

Research Question

- What is the effect of optical exposure (e.g., TV, computer, and video games exposure) on cognitive performances over time, controlling for possible covariates?

Data

- National Longitudinal Survey of Youth (NLSY)
  - Sample size: 9,964, ~50% women, 14-22 year olds in 1979
  - Time span: 1979 ~ 2014
  - For covariates: maternal education, age of first birth, maternal SES, etc.
- NLSY Children (NLSYC)
  - Time span: 1986 ~ 2012
  - Sample size: 11,512, ~49% women, age ranging from newly born to late 20s
  - For outcome variables, predictors, covariates

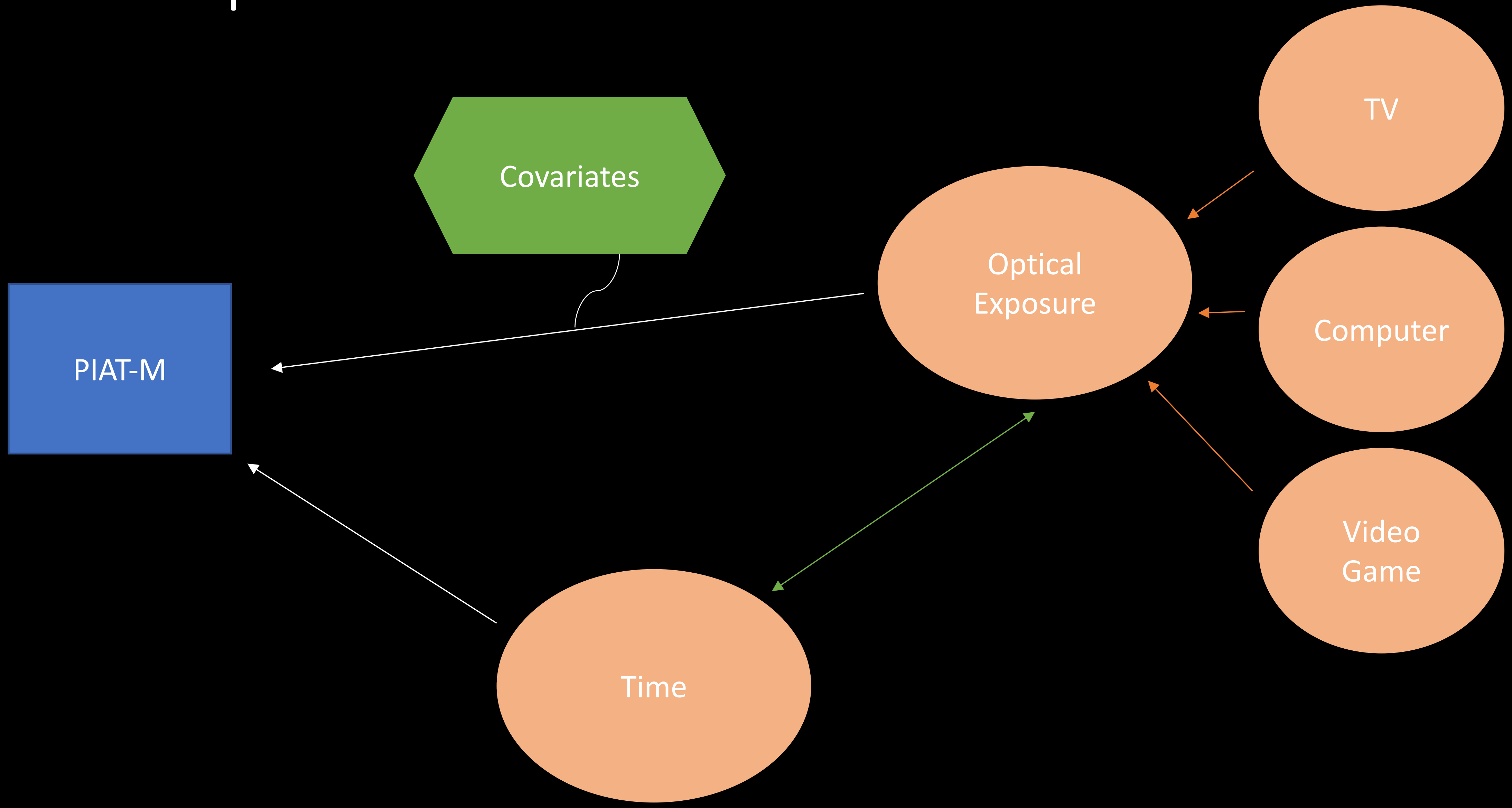
Outcome Variable in NLSYC

- Standardized test scores of the Peabody Individual Achievement Test Math subscale (PIAT-M)
- Maps to fluid intelligence, shows Flynn effect (Rodgers & Wänström, 2007)
- Each survey year 5 ~ 13 yr olds

Proposed Predictors in NLSYC

- TV viewing (13 self-report items from mother and child)
  - # of hrs/day the TV is on in the home
  - whether parents discuss TV programs with child
  - etc.
- Computer usage (8 self-report items from child)
  - # of hrs child uses computer on a typical weekday
  - etc.
- Video game usage (3 self-report item from child)
  - # of hrs child plays video games on a typical weekday, Saturday, and Sunday

Conceptual Model



Hypotheses?

- As optical exposure increases the PIAT-M scores also increase over time.
- TV exposure is relatively the most important type of optical exposure.

Methodology

- Data reduction w/ principle component analysis
  - X - standardized data matrix including the three types of optical exposure var
  - $\mathbf{X'X} = \mathbf{S\Lambda S'}$ → $\mathbf{P} = \mathbf{XS\Lambda}^{-1/2}$

    $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots \mathbf{p}_n]$
- Prediction w/ Multiple linear regression analyses:
  - T = testing variables, C = covariates

$$\mathbf{y} = \beta_0 + \beta_1(\mathbf{p}_1) + \beta_2(\mathbf{t}) + \beta_3(\mathbf{p}_1)(\mathbf{t}) + (covariates) \cdots$$

→

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

$$\mathbf{X} = [\mathbf{T} \cdots \mathbf{C}]$$

References

Dickens WT, Flynn JR (2002). "The IQ Paradox: Still Resolved" (PDF). Psychological Review. 109 (4): 764–771.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. Psychological Bulletin, 95, 29−51.

Greenfield, P. M. (1998). The cultural evaluation of IQ. In Neisser (Ed.), The rising curve: Long-term gains in IQ and related measures (pp. 81−123). Washington, DC: American Psychological Association.

Rodgers, J. L., & Wänström, L. (2007). Identification of a Flynn Effect in the NLSY: Moving from the center to the boundaries. *Intelligence*, *35*(2), 187-196.

Do critical users on online social Q&A communities have broader interests in topics compared to others ?

Jingyuan Zhou

Why?

- Why would certain people attract continuous interest from others in their communities?

- Do they generally have focused but in-depth knowledge on certain topics?
- Or
- Does their broad spread of interests make them interesting?

Related work

- Shah, Chirag, and Jefferey Pomerantz. "Evaluating and predicting answer quality in community QA."
- Kim, Kyung-Sun, Sei-Ching Joanna Sin, and Tien-I. Tsai. "Individual differences in social media use for information seeking." *The Journal of Academic Librarianship* 40.2 (2014): 171-178.
- Jin, Jiahua, et al. "Why users contribute knowledge to online communities: An empirical study of an online social Q&A community." *Information & Management* 52.7 (2015): 840-849.
- Cha, Meeyoung, et al. "Measuring user influence in twitter: The million follower fallacy." *Icwsm* 10.10-17 (2010): 30.
- González-Bailón, Sandra, Ning Wang, and Javier Borge-Holthoefer. "The emergence of roles in large-scale networks of communication." *EPJ Data Science* 3.1 (2014): 32.

How? Data + Computational tools

- Get users that participate in top 1000 questions with highest number of upvotes
- Splitting this network of users into communities with Fastgreedy algorithm
- Within each community, find critical users using Pagerank algorithm
- For each user, find topic distribution of questions they participate in
- Compare finding of critical users and the others

US Covert Operations and Suicide Terrorism

Soo Wan Kim

April 5, 2017

Introduction

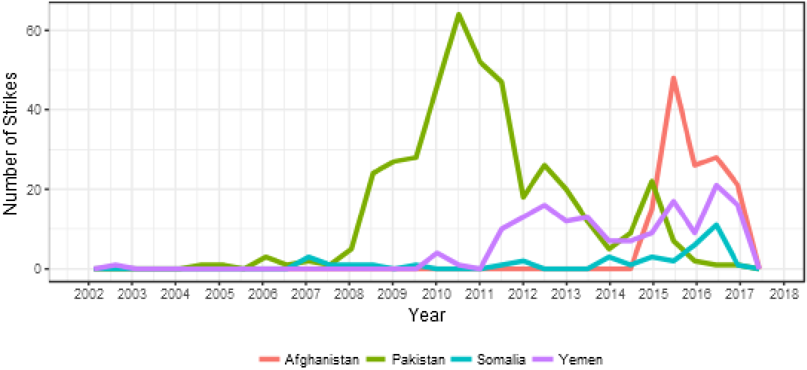
Covert Operation

"An operation that is so planned and executed as to conceal the identity of or permit plausible denial by the sponsor."
(U.S. Department of Defense Dictionary of Military and Associated Terms)

Generally speaking: drone strikes, other air strikes

- The Obama administration relied heavily on drone strikes and other covert operations to target the leadership of terrorist groups
- These operations allow the US to reduce their military footprint overseas, but attract international controversy due to collateral damage to civilians and their secretive nature

Locations of US Covert Operations, 2002 - present



Source: The Bureau of Investigative Journalism (Confirmed attacks only)

Research Question

- Do covert US air strikes delay or hasten subsequent suicide terrorist bombings in the countries where the strikes are carried out?

Theory

- Scholars divided on the utility of air strikes in counter-terrorism
- Rival sets of theories:
  - **Blowback/Backlash**: Strikes anger local populations and increase support for terrorist groups $\rightarrow$ more terrorism
  - **Disruption, degradation, deterrence**: Strikes interfere with terrorist groups' operations, remove key players, and deter would-be terrorists $\rightarrow$ less terrorism

Empirical Findings: Incidence Models

	Scope	US Operations	Terrorist activity	Key Findings
Lyall (2014)	Afghanistan 2006-2011	All air strikes and non-lethal shows of force	Attacks on military targets	**Increase** in incidence of attacks in bombed areas relative to non-bombed
Gill (2015)	Pakistan 2004-2013	Drone strikes	All attacks	**Increase** in incidence and lethality of attacks after drone strike
Johnston & Sarbahi (2016)	Pakistan 2007-2011	Drone strikes	Attacks on civilians	**Decrease** in incidence and lethality of attacks after drone strike

Gaps in the Literature (1)

- Which types of terrorism are relevant? Not all acts are carried out by the groups targeted in US air strikes. Some carried out by rival groups or lone wolves. Impossible to know for certain who committed what. Thus, looking at terrorist activity as a whole may not be helpful for gauging the effect of air strikes on the groups they specifically target.

Gaps in the Literature (1)

- Which types of terrorism are relevant? Not all acts are carried out by the groups targeted in US air strikes. Some carried out by rival groups or lone wolves. Impossible to know for certain who committed what. Thus, looking at terrorist activity as a whole may not be helpful for gauging the effect of air strikes on the groups they specifically target.

Proposed solution

Use suicide bombings only. Suicide bombings are particularly difficult to carry out, require heavy explosives and often specific training. They are the hallmark of professional terrorist cells, not lone wolves. Also, they are widely used by the groups targeted in US air strikes.

Gaps in the Literature (2)

- Do air strikes hasten or delay terrorist activity? Looking at the incidence of attacks across a longer period (e.g. month) does not answer this question because attacks can be spaced closely together or far apart.

Gaps in the Literature (2)

- Do air strikes hasten or delay terrorist activity? Looking at the incidence of attacks across a longer period (e.g. month) does not answer this question because attacks can be spaced closely together or far apart.

Proposed solution

Look at the length of time between attacks.

Question & Hypotheses

- **Research question (restated)**: Do covert US air strikes delay or hasten subsequent suicide terrorist bombings in the countries where the strikes are carried out?
  - **H1 (Blowback/Backlash)**: Terrorist groups behave more aggressively after an air strike and carry out more attacks in quicker succession.
  - **H2 (Disruption/Degradation/Deterrence)**: Terrorist groups are hindered by the effects of air strikes and take longer to prepare and carry out attacks.

Assumptions

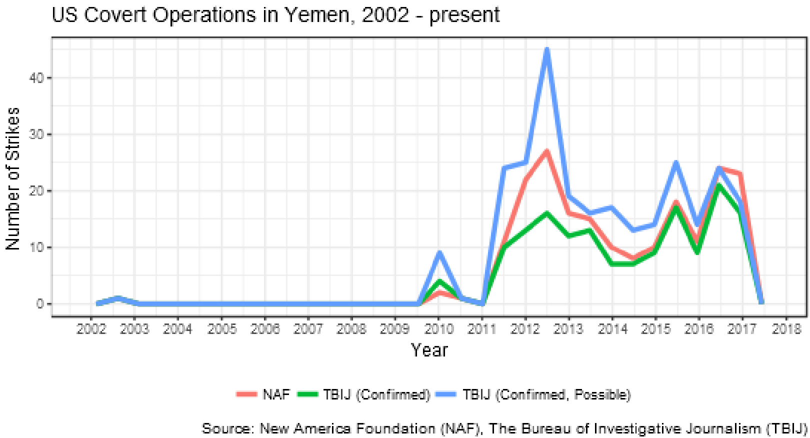
- The effects of drone strikes are not necessarily localized, i.e. terrorist cells may move away from zones targeted by air strikes, or the same terrorist group may respond to an air strike in one part of the country by carrying out an attack in another part of the country. However, they should generally operate in the same country over the short term.
- Terrorist groups do not distinguish between different types of air strikes.

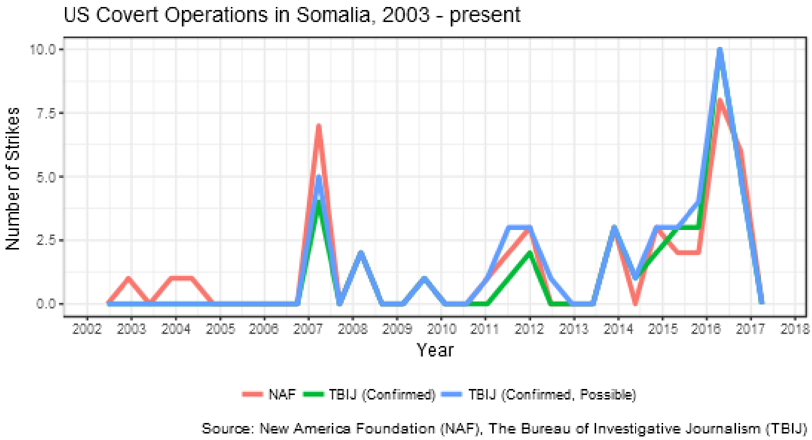
Analysis

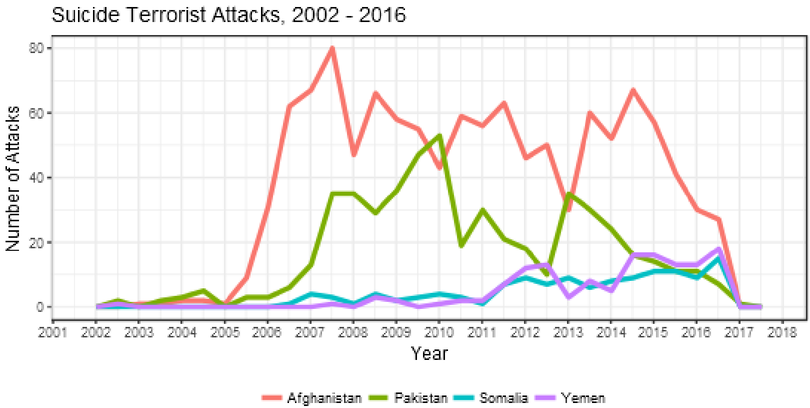
- **Method**: Survival/event history model
- **Unit of analysis**: Country-week
- **Independent variable**: Number of US air strikes (lagged or weighted)
- **Dependent variable**: length of time between individual suicide bombings

Data

- **Countries**: Yemen, Somalia, Pakistan, Afghanistan
- **Years**: 2002-2016
- **Drone strikes data**: The Bureau of Investigative Journalism
  - Independent watchdog
  - Data based on international and domestic media reports, government reports and other accounts
  - Data publicly available for free
  - Only source for air strikes in all four countries
- **Suicide terrorism data**: UChicago's Suicide Attack Database
  - Based on international and domestic media reports
  - Publicly available, free


US Covert Operations in Yemen, 2002 - present
Number of Strikes
0
10
20
30
40
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
Year
NAF
TBIJ (Confirmed)
TBIJ (Confirmed, Possible)
Source: New America Foundation (NAF), The Bureau of Investigative Journalism (TBIJ)


US Covert Operations in Somalia, 2003 - present
Number of Strikes
10.0
7.5
5.0
2.5
0.0
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Year
NAF
TBIJ (Confirmed)
TBIJ (Confirmed, Possible)
Source: New America Foundation (NAF), The Bureau of Investigative Journalism (TBIJ)



Source: Suicide Attack Database

- Gill, Paul. 2015. "The Impact of Drone Attacks on Terrorism: The Case of Pakistan." London: Remote Control Project. http://remotecontrolproject.org/wp-content/uploads/2015/06/Paul_Gill_drones_terrorism_Pakistan.p
- Johnston, Patrick B., and Anoop K. Sarbahi. 2016. "The Impact of US Drone Strikes on Terrorism in Pakistan." International Studies Quarterly 0: 1–17. doi:10.1093/isq/sqv004.
- Lyall, Jason. 2014. "Bombing to Lose? Airpower and the Dynamics of Violence in Counterinsurgency Wars." http://dx.doi.org/10.2139/ssrn.2422170.