

Student: Maya Bhat-Gregerson  
Mentor: Mohammed Zakaria  
Date: August 11, 2018  
Course: Springboard - Introduction to Data Science

Ideas for capstone projects

## **(1) Predictive value of Tweets indicating suicidal ideation and depression and correlation with completed suicides**

### *The problem*

In public health surveillance, typically epidemiologists look at historical data to understand trends in diseases or conditions that lead to death. One common measure is to calculate a mortality rate for a given health concern but the final data set will include deaths that occurred from 6 to 18 months ago. Alternatively, some epidemiologists look at a weekly feed of provisional death data to track the count of deaths due to a particular disease or event. In either case, we rely heavily on lagging indicators and can only paint a historical picture of events that have already happened. With the epidemiological tool currently in use we don't have a way to:

- (a) foresee changes in patterns (magnitude) of mortality due to a given disease or event, or
- (b) see if a particular community-wide public health intervention has made a noticeable impact in reducing the problem.

### *Proposal*

To see if there is a predictive relationship between tweets expressing suicidal sentiments and completed suicides by looking at historical data.

### *Why the problem is important and who might care about a potential solution?*

Mental health issues are among the most under-recognized health problems we face. In part, this is because we have few ways to measure the scope of the issues like depression and suicide ideation in our communities. If it is possible to find a way to predict trends in completed suicides by monitoring the volume of tweets expressing suicide ideation it may allow public health and mental health agencies to quantify the magnitude of the problem in our state and potentially implement population level interventions to prevent these deaths. It may also allow us to track the effect of public health campaigns to prevent suicide.

### *Link to background paper(s)*

<https://www.sciencedirect.com/science/article/pii/S2468696417300605>

<https://www.sciencedirect.com/science/article/pii/S2214782915000160#s0020>

## **(2) Predictive value of ambient temperature, air quality, and built environment factors on risk of death due to heat-related causes**

### *The problem*

The problem is the same as state in idea (1) except in this case the deaths of concern are heat-related. For the past few years we have experienced unusually hot summers (more days with

higher than average temperatures). In some years the high temperatures have been accompanied by poor air quality due to wildfires in various parts of the Pacific Northwest. We are anticipating an increase in heat-related deaths over the coming years as the trend in elevated summer temperatures continues. Again, we currently rely on historical data to understand the characteristics and circumstances of persons who died of heat-related causes.

#### *Proposal*

To examine heat-related deaths that occurred in the past 10 years in Washington State and look at the weather, built environment, and social factors associated with the deaths to see if there's any potential for predicting risk of heat-related deaths in future years.

#### *Why the problem is important and who might care about a potential solution?*

Public health agencies would be interested in tools that can help them predictive relatively accurately what populations are at higher risk of succumbing to heat-related stress. These deaths may be preventable with timely response.

Factors that might contribute to higher risk of these deaths include temperature, air quality (particulate matter), census tract information for the place of death of each decedent such as average income, age of homes, population density etc. If there's any information available on the percent of a census tract covered by tree canopy or green space that would be an interesting factor to look at as well.

#### *Link to background paper(s)*

<https://jech.bmj.com/content/jech/64/9/753.full.pdf>

<https://health2016.globalchange.gov/temperature-related-death-and-illness>

### **(3) Identifying homelessness in Washington State deaths**

#### *The problem*

In the current death registration process in Washington State's there is no systematic way for medical examiners/coroners or funeral homes to report that the decedent was homeless. Without this information, it is virtually impossible to quantify accurately the health problems that cause premature mortality among the homeless in our state. The medical examiner/coroner filing the death certificate may enter a word or a phrase to indicate homelessness in one of the free text fields in our electronic death registration system. There is no consistency in whether and how homelessness is reported.

#### *Proposal*

To analyze text fields in Washington State death data to classify decedents as having a permanent residence at the time of death or being homeless. I have obtained from King County Medical Examiner's office (county in which Seattle is located) 10 years worth of names and death certificate numbers of King County residents deemed by the medical examiner to be homeless. Although not all county medical examiners or coroners follow the same method of ascertaining homelessness among decedents, it is a starting point to trying to identify homelessness in the rest of the state.

*Why the problem is important and who might care about a potential solution?*

Homelessness is a growing problem across the country but we know very little about the health problems that affect this population. Most attempts to understand the health of this population is limited to work undertaken by individual homeless shelters. In order for public health and social service agencies to provide appropriate services that are accessible by homeless individuals it would be useful to understand the leading causes of death in this population as well as some basic demographic information.

*Links to background information*

<https://www.pressreader.com/usa/san-francisco-chronicle/20180311/281513636669274>

<https://www.kingcounty.gov/depts/health/locations/homeless-health/healthcare-for-the-homeless/~media/depts/health/homeless-health/healthcare-for-the-homeless/documents/medical-examiner-analysis-homeless-deaths.ashx>

#### **(4) Linking death certificates with hospitalization discharge records**

*Problem*

Beyond what is reported through our electronic death registration system we don't have any health history on decedents in Washington State. Separately, we collect hospital discharge data that includes up to 20 diagnosis codes for billing purposes associated with each hospital inpatient visit. Currently, we do not have a systematic way to link the two data sets. Without this linkage we do not have a systematic way of assessing the quality of cause of death information submitted on death certificates.

*Proposal*

To develop a way to match hospital discharge data to death certificate data that accounts for name changes or variations, changes in address, etc.

*Why the problem is important and who might care about a potential solution?*

Without a way to link death certificates to any prior hospitalizations we are missing an opportunity to bring together information that can create a more complete picture of the health events leading to death. Linkage of the two data sets would allow comparison of the main diagnoses codes of persons who were hospitalized with the underlying cause of death reported on the person's death certificate to check the accuracy of the latter. It would also allow other types of analysis e.g. survival analysis of patients undergoing different types of procedures in hospital to address the same health problem;

<https://academic.oup.com/ije/article/29/3/495/771321>

[https://www.cdc.gov/pcd/issues/2015/14\\_0299.htm](https://www.cdc.gov/pcd/issues/2015/14_0299.htm)

<https://www.futuremedicine.com/doi/full/10.2217/cer.13.4>

<https://www.ncbi.nlm.nih.gov/pubmed/15019007>