CS171 Detailed Project Plan

Lennon Bensaou and Andrew Zhou

November 7, 2016

1 Basic Info

We are Andrew Zhou and Lennon Bensaou (Team "Benzhou"), two Harvard College seniors taking CS171 in Fall 2016. Our working project title is "The Forgotten Epidemic: Exploring America's Misunderstood Mental Health Crisis." We can be contacted at andrewzhou@college.harvard.edu and lennonbensaou@college.harvard.edu, respectively.

2 Background and Motivation

Both of us are keenly interested in the topic of mental illness. Lennon is a Psychology concentrator and took Psych 18: Abnormal Psychology the year before last, and so is well-versed in the science of mental illness: its conjectured causes; its signs, symptoms, and prognoses; its effects on sufferers' lives; our current scientific understanding of just what "mental illness" is; and the best ways to treate it that we know of. Andrew a Computer Science concentrator, and in addition to being interested in the hard data (both sociological and scientific) concerning mental illness, is committed to the social justice issues that are inseparable from mental illness, such as its associations with ethnicity, gender, socioeconomic status, sexuality, and other demographic factors. We are also both interested in questions of mental health awareness and advocacy, particularly societal perceptions of these issues: how have these perceptions evolved over time, and what is society's current understanding of mental illness and the mentally ill? These are topics that we would like to explore, and we are excited about analyzing and presenting the intricacies of the problem through our visualizations.

3 Project Tasks

We have outlined seven major tasks that we would like to tackle in our project. This is a very ambitious goal, and we are unsure that we will address all of them with the detail they deserve, but we are using this list as a starting point. These tasks are:

1) Introduction and Scientific Background
2) Demographic Information
3) Symptoms, Causes, and Other Associations
4) Treatments

5) Comorbidity
6) Social Understandings and Perceptions of Mental Illness
7) Mental Health Resources

We will address each of these in turn below, and we will outline our ultimate goals for the project in Section 4.

3.1 Introduction and Scientific Background

As part of our storytelling approach, we would like to educate the general public on the general subject of mental health. Therefore, we will provide an overview of the topic as well as information concerning researchers' current scientific understanding of "mental health" and its conjectured causes as well as signs, symptoms, and prognoses. We will provide brief overviews of the mental illnesses we plan to focus on—major depressive disorder, bipolar disorder, anxiety disorders, schizophrenia, and ADHD—as well as a brief history of societal and scientific understandings of mental illness. It is our hope that theis lead-in will serve as an introduction to our storytelling approach, and that it will provide the necessary context to make sense of our visualizations and in-depth explanations that will follow.

3.2 Demographic Information

The first major area of mental illness that we would like to explore is its demographic associations. Owing to a confluence of circumstances, mental illness strikes different demographic groups at different rates, and we are interested in the precise nature of these differences. For instance, we might display the incidence of schizophrenia in men versus women, or in heterosexual versus LGBT populations.

We wish to take note of both the most and least vulnerable groups, noting which mental illnesses they suffer from and at what rates, as doing so may help us understand why some groups are particularly predisposed to mental illness while others are more "resilient." We will highlight particularly notable data concerning the most and least vulnerable populations, and will consult and cite the literature of the field to see how sociologists and scientists explain these disparities.

Some specific demographic factors we are interested in are: ethnicity; gender; sexuality; religion; socioeconomic status; level of education; history of military service; and geographical location. This last factor may be broken down into various sub-factors, such as state of residence; urban versus rural living; polluted versus clean air; population density; and any other geographical traits that we find are important. We plan to present these data in both a choropleth and a bar chart, which are tentatively illustrated in Section 6.

3.3 Symptoms, Conjectured Causes, and Other Associations

In addition to demographic factors, we would like to explore the association of various mental illnesses with certain symptoms, potential "causes," and various other associations, which we will define later in this section. For example, we might explore the incidence of self-harm among individuals with bipolar disorder, or the association of anxiety disorders with Body Mass Index (BMI).

We will note that it is somewhat arbitrary for us to distinguish these factors from "demographic" ones, but we are qualitatively separating these categories as demographic factors are associated with "populations" whereas these factors are not. For instance, we consider Asian Americans a population, but do not consider those who self-harm a population.

It is very difficult to distinguish between a symptom of mental illness, a cause of mental illness, and a factor that is correlated but not causatively associated with mental illnesses, so we group these factors in the category of "Assocations." Some associations we are interested in are: various health metrics, e.g. history of heart disease, history of stroke, and BMI; history of self-harm and suicide attempts; history of drug use; employment, e.g. whether an individual is employed, and if so, whether this individual has missed work due to their illness; and homelessness. We plan to present these data in a bar chart, tentatively illustrated in Section 6.

3.4 Treatments

We would also like to explore treatments received by mental health patients. We distinguish this section from the above two because we would specifically like to explore the efficacy as well as incidence of such treatments. Such treatments include the use of prescription medication; history of seeing a medical professional; history of seeking "alternative" treatment, e.g. homeopathic or spiritual approaches; history of inpatient hospitalization; and history of outpatient care. We will examine both the incidence of particular treatments, such as demonstrating what percent of bipolar individuals take medication, as well as the efficacy of these treatments, such as what percent of individuals seeking "alternative" treatment report improved symptoms.

3.5 Comorbidity

An important association we would like to visualize separately and specifically is comorbidity. Comorbidity refers to the condition of suffering from two or more illnesses simultaneously, and we are interested specifically in the comorbidity of mental health conditions. Patients who suffer from two or more mental illnesses are a particular vulnerable population, and we would like to examine the incidence of comorbidity as well as the incidence of specific combinations of illnesses, e.g. what percentage of mentally ill individuals suffer from two comorbid conditions, and what percent of bipolar individuals also suffer from schizophrenia?

We plan to present data on the incidence of comorbidity in a pie chart, tentatively illustrated in Section 6.

3.6 Social Understandings and Perceptions of Mental Illness

We are particularly excited about gathering and visualization data concerning the social understandings and perceptions of mental illness in American society. We believe that awareness is crucial to improving the plight of the mentally ill in the United States, as ill individuals will be more likely to seek treatment if these issues are more discussed and less stigmatized. We have not included sketches for visualizing these data, as we are still considering our approach, but we will outline several ideas we have.

One approach we are considering (and particular excited about) is using the Google Ngram Viewer to track the incidence of certain mental-health related terms in Google's text corpora over time. By doing so, we can see how society's understanding of mental health has shifted over the years based on its vocabulary for mental illness. For instance, scientists used to use the term "hysteria" to describe women with fluctuating moods, but that misogynistic term has been discarded in the mental health profession. Other shifts include the renaming of "manic-depression" to "bipolar disorder" and the abandoning of the term "asylum" for mental health facilities over time. We believe that visualization these data may be done in particularly innovative ways, and will consult with the course staff as we develop more ideas.

In addition to tracking the historical evolution of our understanding of mental health, we would also like to demonstrate what beliefs present-day individuals have about mental health versus the reality of the situation. For instance, what percent of individuals consider the mentally ill potentially violent, and what percent of mentally ill people commit violent actions? Is there a gap between perceived rates of mental illness in society and actual rates? We will note that it may be difficult to get specific data on these questions, so we will adjust our plans for this section as we do more research.

3.7 Mental Health Resources

As a bookend for our project, we would like to include links to various mental health resources. This section is perhaps tangential to our goal of visualizing data, but it is in line with our goals for the project (see 4) and seems obligatory given the topic we have chosen. A potential resource, for instance, would be the Behavioral Health Treatment Services Locator created by the Substance Abuse and Mental Health Service Administration (SAMHSA). We will be sure to select these resources carefully, as linking to nonreputable sources would be a very grave error.

4 End Goals

Our end goals are fairly simple. Our first goal is to raise awareness of mental health issues afflicting Americans. We hope that by displaying data concerning the prevalence and distribution of mental illness, we can help the general public understand the nature and extent of mental health issues. We would also like to draw attention to demographic groups that are particularly susceptible to mental illness, so as to better inform the public of how factors such as ethnicity, gender, and sexuality are associated with mental illness. We would also like to illustrate the gravity of the crisis, such as by indicating the prevalence of self-harm, suicide attempts, unemployment, and homelessness among the mentally ill.

Our second goal is to change the general public's perceptions of mental illness and decrease the societal stigma against mental illness. There exist many prejudicial stereotypes about the mentally ill: they are all "crazy"; they are violent; they are simply faking it. We hope that by presenting a wealth of data concerning mental illness, we can overcome stereotypes and misunderstandings and better inform the general public of this health crisis. Hopefully by doing so, we can also encourage those who are resistant to seeking treatment due to the stigma to do so. We provide a link to mental health resources as additional motivation to learn more and perhaps explore the option

of seeking treatment.

Our third goal is to provide scientific and sociological information about the history of both mental illness and our understanding of it. This goal is part of our storytelling approach: we would like to educate viewers of our project of how both scientific and societal understandings of mental illness have changed over time. We believe that if viewers are better informed as to the historical aspects of mental illness, they will understand the crisis better and be less liable to stereotype the mentally ill.

5 Data Acquisition and Processing

5.1 Acquisition

We have begun to look into various sources of mental health data, and have some very promising leads. The Substance Abuse and Mental Health Service Administration (SAMHSA) conducts an annual survey called the National Survey on Drug Use and Health (NSDUH), and we were able to locate the survey results from 2005 to 2014. This survey asks a representative sample of American citizens about their drug use habits and mental health issues the year prior, and will give us a snapshot of mental health issues in the US at any time in the last decade. Additionally, SAMHSA conducts an annual survey of mental health institutions called the National Mental Health Services Survey (NMHSS), the 2010 version of which we have located. The NMHSS surveys mental health institutions yearly on their admission statistics and the treatment types they offer, and will give us an idea of hospitalization statistics and the different kinds of treatments applied to mental health. As we have not been able to locate the results of all surveys conducted by SAMHSS, we plan to contact the organization regarding access to their data.

Though dated, the National Comorbidity Survey (NCS) is the most comprehensive dataset we have located concerning mental health. The baseline NCS was conducted in 1990-2 and then its respondents were reinterviewed in 2001-2 for the NCS-R. The dataset is an extraordinarily detailed survey of respondents, and asked them questions concerning a wide range of mental illnesses, psychiatric drugs, life experiences, demographic circumstances, and many other factors.

As mentioned previously, we plan to use the Google Ngram Viewer to trace the development of mental health vocabulary over time. Also, Lennon took Psychology 18: Abnormal Psychology the year before last, and we have contacted his professor for the class regarding potential datasets we could look into. We are also planning on reaching out to the Harvard Libraries and the Psychology department concerning the acquisition of data.

5.2 Processing

Some of our datasets are too large, so we have begun to process them. In particular, the NSDUH for 2014 is a 500MB .tsv file, far too large to load in d3. We wrote a python script to filter out extraneous variables, and after doing so were able to load the resulting file. We also plan to filter the NMHSS, baseline NCS, and NCS-R so they are easier to work with.

As for the Google Ngram Viewer, we are only interested in a select subset of mental-health related terms (e.g. "hysteria," "asylum," and "manic-depressive"), so we may be able to manually query the database for the terms of interest. We may write a script to query the database if we decide eventually to work with a very large number of terms.

6 Sketches

We provide sketches of our demographic/association-based choropleth and bar chart and our comorbidity pie chart in Figures 1 and 2. (Figure 1 is 2 sketches.) Note that some of our images are displaced due to the typesetting environment. Note that the full-resolution scanned images can also be found in the .zip file.

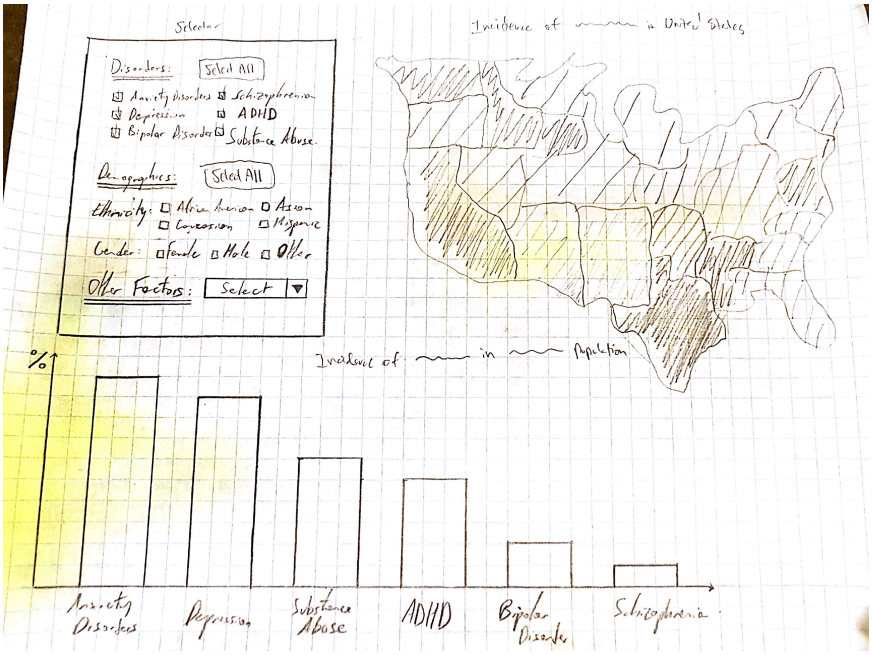

Figure 1: Bar Chart and Choropleth of Mental Illness by Demographic and Association

7 Interaction Storyboard

We provide a sketch of our interaction storyboard for the comorbidity pie chart in Figure 2. Note that some of our images are displaced due to the typesetting environment. Note that the full-resolution scanned images can also be found in the .zip file.

8 Webpage Layout and Storytelling

We provide a sketch of our webpage layout and storytelling in Figure 4. Note that some of our images are displaced due to the typesetting environment. Note that the full-resolution scanned

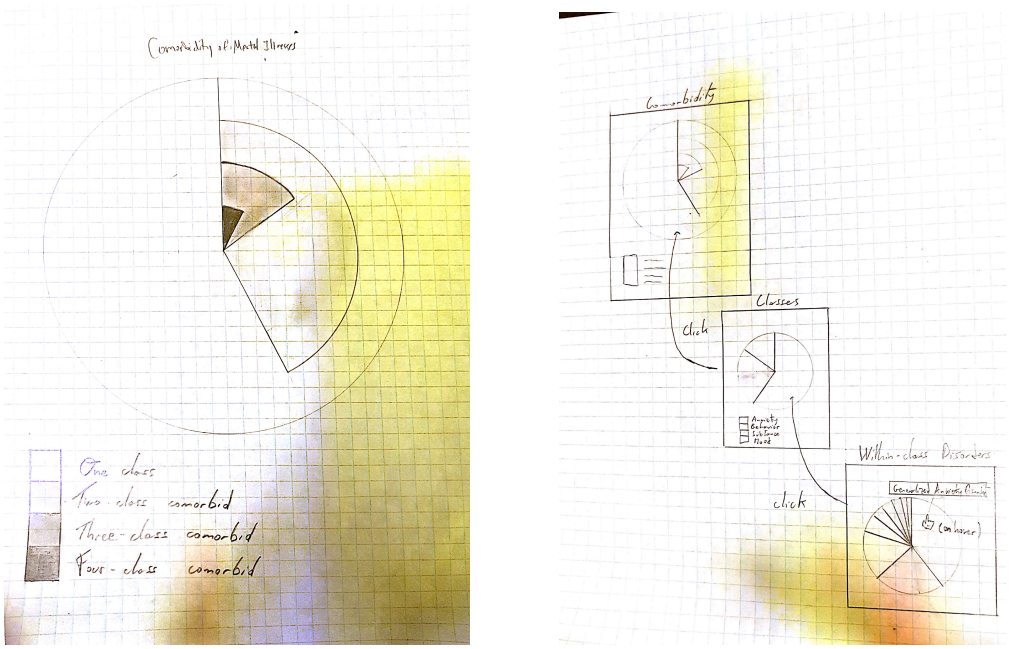

Figure 2: Comorbidity Pie Chart (left) and Pie Chart Interaction (right)

images can also be found in the .zip file.

9 Project Timeline

We plan on meeting every Friday from 2-5pm to work together and check our progress. We will set deadlines for Fridays.

- November 11: Found data for the choropleth, bar chart, and pie chart. Contacted SAMHSA and followed up with Lennon's professor for assistance. Have a rudimentary website framework up. Have preliminary choropleth, bar chart, and pie chart up and running (without selector/interactivity).
- November 18: Found Google Ngram data. Have selector working for choropleth and bar chart and interactivity working for pie chart. Compiled external mental health resources. Implemented informative text-boxes. Begun writing text for website. Compiled list of references.
- November 25: Implemented the Google Ngram visualization. Implemented informative text-boxes. Begun work on video. Finished writing text for website. Begun compiling and editing process book.
- December 2: Continued work on video. Begun implementing "Optional Features" if we have time.
- December 9: Finished process book. Finished video. Finished "Optional Features." Compiled and edited README file. Last-minute fixes.

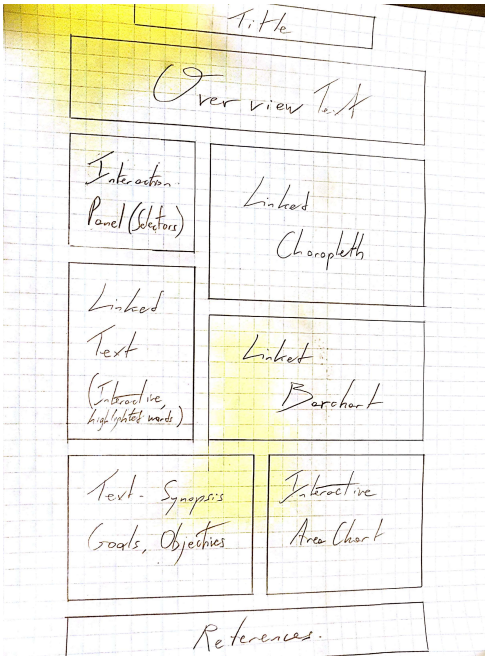

Figure 3: Website Layout and Storytelling

10 Feature List

10.1 Must-Have Features

These are specific features that we are committed to having.

- Choropleth displaying incidence of mental illness by geography
- Bar chart displaying incidence of mental illnesses
- Selector connected to both of the above to filter by demographic and association
- Interactive comorbidity pie chart
- Informative introduction/overview
- Text-boxes highlighting and explaining particularly important information
- Video introduction and walkthrough
- Links to vetted external mental health resources
- List of references

10.2 Good-To-Have Features

These are more general ideas whose visualization types we have not precisely planned out.

- Timeline of the changes in mental health terminology over time based on Google Ngram
- Visualization indicating current public attitudes toward mental illness
- Visualization indicating the efficacy of drugs and treatments

- Annotations that, when clicked, set the selector to a particularly important demographic/association and provide an explanatory text-box, e.g. "Asian American Veterans are more likely to suffer from..."
- Explanation of the scientific underpinnings behind various mental illnesses

10.3 Optional Features

These are very vague ideas that we may tackle if we have extra time.

- Map of treatment facilities in America, along with treatment types they offer
- Timeline of the evolution of our scientific understanding of mental health, e.g. "Lobotomies were first performed in..."
- Animating the timeline(s) to display events one by one when a "play" button is clicked
- Visualization of costs of medical treatment for mental illnesses

11 Team Roles

As mentioned on the Team Member Agreement, Andrew plans on working principally on Data Wranging and Implementation, while Lennon will focus on Targetting and Design. Both will share the task of Evaluation and searching for the data. Additionally, Andrew plans to oversee the writing of the project book, while Lennon plans to write the text of the website, particularly the scientific explanations.