

CSCE 5320 Project Description (Story Telling)

The overall storyline focuses on “5W + H”: Who is it about? What happened? When did it take place? Where did it take place? Why did it happen? How did it happen.

“Chapter 1 Life” should accurately and passionately describe real-world problems facing the targeted users or audience:

- **Who** are the people or communities in need of help?
- **What** problem happened to them?
- **When** did the problem take place?
- **Where** means two things: 1) The environment and settings that the people or the community is living in, and 2) the place/location where the problem take place.
- **Why** means the possible causes and/or origin of the problem?
- **How:** If you would like, you can add a dimension of how. How did it happen? Sometimes, the answer to how can be covered by what, when and where.

Chapter 2 Data. The objective of Chapter 2 is to use stories as hands-on practice to link large-scale data sets with real-world problems identified in Assignment 1. Meanwhile, this is also a process for data cleaning, data preprocessing, and data management using the tools learned in this course.

This should guide the process of identifying data sets that are most relevant to the real-world problem in Assignment 1:

- **Who** is the data set about? **Who** were sampled in this data set? **Who** were over sampled or under sampled? Are they representative of the main characters in Assignment 1? Is there any identifiable information or is there any risk of disclose identifiable information? This is fundamentally about the sampling issue, and anonymity.
- **What** events, activities, behaviors, and observations etc. are recorded by the data set? Does the data set record the targeted events, activities, behaviors, etc. in Assignment 1? This is fundamentally about the variables.
- **When** did the event, activity, behavior, and observation, etc. take place? **When** were the data collected? Is it longitudinal or cross-sectional? Are they real time data? How old or fresh are the data? To what extent generalization can be made across time to inform Assignment 1? This is fundamentally about the temporal structure of the data set, and the external validity of the data set across time.
- **Where** did the event, activity, behavior, and observation, etc. take place? **Where** were the data collected if the information is available? What does the geographical coverage of the data set look like? Does the data set contain geographical information (GIS)? Is this a local, regional, national, or global data set? To what extent generalization can be made across settings to inform Assignment 1? This is fundamentally about geographic variables in the data set, and the external validity of the data set across settings.

- **Why** did the event, activity, behavior, or observation etc. take place? **Why** were the data collected?
- **How:** If you would like, you can add a dimension of how. How did it happen? Sometimes, the answer to how can be covered by what, when and where.

“Chapter 3 Users” is a story about the user using the application/visualization:

- **Who:** the main character is the targeted user or audience?
- **What** can the application do? What does the visualization show?
- **When** can the user use the application/visualization?
- **Where** will the visualization and applications be deployed, for example, mobile phones, the web etc.?
- **Why** is the visualization or application useful to the user?
- **How** will the people/the community use this application or visualization to make changes?