# Project Assignment – Worksheet 1: Outline

## Background

For this project, you are the researcher who has just finished their data accumulation phase! You have so much data and don’t know what do to! Thankfully you took BINF5503 and can apply the tools we’ve learned this semester to look through your data, conduct analysis and prepare a poster for a conference.

In this assignment, you will form groups of 2-4 and will select one dataset from a group of provided datasets that aligns with your group’s interest (see the Datasets folder on Blackboard). Using your chosen dataset, your group will test a novel research question, complete a self-directed analysis, and tell the complete data story using a poster.

There are four parts to the project:

* **Outline (14%)**
* Draft Poster (14%)
* Presentation (17%)
* Final Poster (17%)

For this first outline, identify a dataset you would like to work with, conduct a preliminary exploration of the data, propose a research question(s), and discuss next steps.

## Objectives

* Understand and explain the context of your dataset using the materials included and through additional research (What is the gap in the field? How was the data collected? What are the variables of interest?).
* Propose at least one novel research question, a hypothesis, and a prediction based on the research previously conducted
* Select several statistical analyses that can be conducted to test your research question, and justify how the analysis will contribute to understanding the dataset/field
* Propose visualizations and describe the packages and functions you would need to create them
* Highlight at least one potential roadblock or area of concern and how you plan to address it

## How to complete this assignment

Submit a document that contains:

* A brief introduction of the field your data comes from (i.e. what did your light research tell you?)
* Your research question(s) and corresponding hypotheses and predictions
* A description of your data alongside exploratory code investigating your data
* How you and your group aim to answer your research questions (propose statistical analyses, visualizations and a solution to a possible roadblock)

This document can be in either a Word or RMarkdown (RMD) document with exploratory code and tables/ figures (if applicable) to support your research question(s).

Make sure to run code sequentially from start to end and knit your code consistently.

## Submission materials

Submit the following file(s) to the correct assignment on Blackboard before the due date:

* Research document (eg: PDF file, exported from Word or RMD)
* Code (eg: R or RMD file)

**This assignment will be due on Friday June 27th at 8pm ET.**

## Course grade

This outline is worth 14% of the final course grade

## Support

You are recommended to start the assignment as soon as possible.

Questions should be posted on the Blackboard discussion board under the Worksheet 1 discussion so that everyone has the same information and to reduce repeated questions.

You are permitted to post incomplete problematic code in your question – also include any errors that you encounter or an explanation of why the code is not behaving as you are expecting. While you can use screenshots for code and errors, it would be better to copy and paste text for easier searchability and reproducibility.

You are not permitted to post completed code on the discussion board or otherwise share completed code with others in the course or other platforms. When responding to your peer’s posts on the discussion board, do not respond with completed code but instead respond with suggestions such as “please check the name of your object for typos”.

## Rubric

|  |  |  |  |
| --- | --- | --- | --- |
| Evaluation | Does not meet expectations | Meets expectations | Exceeds expectations |
| Exploration of dataset and topic  10 pts | Exploratory code partially explores the structure of the data, including an overview of the rows and columns  Data wrangling proposed to prepare datasets for further analysis  No background of the dataset or field provided. No citations    0-3 pts | Exploratory code begins to explore the structure of the data, including an overview of the rows and columns  Data wrangling proposed and partially completed to prepare datasets for further analysis  Some background of the dataset and/or field provided. 1-2 citations.    4-6 pts | Exploratory code thoroughly explores the structure of the data, including an overview of the rows and columns  Data wrangling proposed and partially completed to prepare datasets for further analysis  Background of the dataset and field provided that showcases a clear understanding of the topic. 3+ citations    7-10 pts |
| Novel research question and hypotheses    15 pts | Research question, hypotheses, and predictions are formatted and structured incorrectly  Research question cannot be answered using the dataset  Predictions are not justified    0-5 pts | Research question, hypotheses, and predictions are formatted and structured mostly correctly  Research question can be answered using the dataset  Predictions are justified using prior knowledge  6-10 pts | Research question, hypotheses, and predictions are formatted and structured correctly  Research question is novel, clear, and applicable, and can be answered using the dataset  Predictions are justified using prior knowledge and research  11-15 pts |
| Explain analysis proposal      25 pts | Identify a comparison in the dataset    Propose some steps for the analysis. Direction is unclear  Identify some relevant functions and packages but key tools are missing    0-7 pts | Analysis addresses novel research question    Propose steps required in the analysis. Consider some areas of concern    Identify some relevant functions and packages that will be relevant to your analysis    8-16pts | Analysis addresses novel research question(s)    Propose key steps in the analysis that will be required and identify areas you anticipate may cause concern  Clearly identify all functions and packages that will be relevant to your analysis    17-25 pts |
| Propose visualizations  25 pts | Propose some visualizations but most are inappropriate for data and/or analysis  Errors in visualizations.    Describe or identify some relevant functions and packages but key tools are missing    0-7 pts | Propose some visualizations for data and analysis    Describe or identify some relevant functions and packages that will be relevant    8-16pts | Clearly propose visualizations for data that are appropriate for the analysis    Clearly describe or identify all functions and packages that are relevant    17-25 pts |
| Anticipated roadblocks    10 pts | Show consideration for limitations of the study    0-3 pts | Identify an anticipated roadblock, and suggest how it can impact the analysis    Propose an action to minimize or address the roadblock  4-6 pts | Identify an anticipated roadblock and thoroughly explain why this may be problematic for your analysis    Propose actions that can be used to minimize or address the roadblock  7-10 pts |
| Efficiency of design    10 pts | Code is written with functions that attempt to address the task. Errors are present. Writing is not succinct    0-4 pts | Code is written with reasonably suited functions for the tasks. No errors present and with mostly efficient code design    5-7 pts | No errors present in code and code accomplishes the task with excellent code efficiency.    8-10 pts |
| Formatting    5 pts | Code is missing a large portion of formatting    0-1 pts | Code is mostly annotated or over-annotated such that it is sufficient to assist the reader with following the methodology and rationale of the analysis    2-3 pts | Code is well-annotated but not over-annotated making it easy for the reader to follow the methodology and rationale of the analysis      4-5 pts |