

# CWRU DSCI-451: SemProj-BasicPremise

*Roger H. French, JiQi Liu*

*20 September, 2018*

## Contents

<b>2 DSCI451: Data Science Semester Research Project</b>	<b>1</b>
2.1 Type of project . . . . .	1
2.2 Purpose of Project . . . . .	1
2.3 Basic Steps Involved . . . . .	1
2.4 Variables . . . . .	2
2.5 Packages . . . . .	2
2.6 Timeline . . . . .	2
2.7 References . . . . .	2
2.8 How to make your report . . . . .	2

## 2 DSCI451: Data Science Semester Research Project

A Data Analysis/Prediction & Modeling Project

### 2.1 Type of project

Is it

- Data Analysis of a defined dataset
- building a R package, or
- building a Python package
- other (define)

### 2.2 Purpose of Project

Please provide a short description of the project,

- its background,
- technical approaches you'll use
- and final output.

Similar to an abstract in a scientific paper

### 2.3 Basic Steps Involved

Steps needed to complete the project. For example

- Defining the ideal dataset
- Steps in cleaning and tidying the data, etc.

## 2.4 Variables

Short description of all the variables involved and

- how they are related to the purpose of the project.
- This will become Databook with time.

## 2.5 Packages

List of packages that you are most likely to use

- with their associated functions.
  - things like ggplot (R),
  - or Sci-kit Image (python)

## 2.6 Timeline

Rough timeline to achieve the steps described above.

You may want to make Gantt Chart to present the timeline.

## 2.7 References

Associated references

## 2.8 How to make your report

The report is done as an Rmarkdown document,

- which can be run/compiled to produce two versions of the report as a pdf.
  - One shows your R code and figures,
  - and the other doesn't show R code, just your figures.

You'll then turn in a zip file (and leave a copy in your repo),

- with the dataset
  - (if its not to huge, if it is large, can you make a smaller dataset),
- Rmd file that works,
- and the two pdf reports.

Just will do a pdf report to turn in for

- 1st Report Out
- 2nd Report Out
- 3rd Report Out

You will make .Rpres slides for

- 1st Report Out
- 2nd Report Out
- 3rd Report Out

The license choice of CC-BY-SA 4.0 is suggested

- so that others can use and build on your codes, in an open-source manner.

More restrictive licenses,

- will be needed if your dataset comes from a CWRU funded research project.