Final Individual Project Instructions

Each student is required to write a R Markdown PDF report for analyzing the COVID-19 data listed below. Your report must provide the sections requested in the grading rubric below. You will need to import and tidy the data, and then come up with your own research questions that you would like to answer from the data by performing exploratory data analysis. If you would like to perform a predictive model to answer your research questions that is fine but it is not required. Try to be creative in your analysis and investigate the data in a way that your classmates most likely will not.

Your report is expected to tell a story with the data. Do not just report some statistics that you find but, rather, to provide a coherent narrative of your findings. Some examples of the type of report are as follows.
https://rpubs.com/bradleyboehmke/final_project_example
https://rpubs.com/murphm6/233993
https://rpubs.com/dyang9411/301598
Note that the above examples are presented in HTML format but you are required to submit the PDF format of your own report.

Your final individual project must focus on at least one of the following datasets available at the Resources tool on the course's Brightspace site:
- us.csv
- us-states.csv
- us-counties.csv

The above three datasets contain the US COVID-19 data till 01/11/2022, and were originally obtained from The New York Times github repository:
https://github.com/nytimes/covid-19-data
See the github repository for the detailed data descriptions. Note that the github repository tracks COVID-19 with daily updates. *Thus, you are required to download the above three datasets from the Brightspace not the github repository*. You may bring in some additional data elsewhere to improve the analysis on the required data, but you need first try a submission with the additional dataset files to the Brightspace to ensure that the files are not too large to upload. Note that you are allowed to resubmit your final project unlimited times before the deadline.

You are required to submit the R Markdown file (.Rmd), its generated PDF output file (.pdf), and any other files (except the above 3 CSV files), which will enable me to reproduce your report by knitting your .Rmd file. Note that any missing file will result in a zero grade of your final project. Your final individual project will be temporally graded on a 50-point scale according to the grading rubric below, but will be eventually counted into your final grade with the weight as per the syllabus. Your final project must be submitted to Brightspace by 11:59 PM (US eastern time) on January 28, 2022. Late submissions will receive a grade of zero.

Section	Standard	Possible Points
Introduction/ Synopsis	**1.1** Provide an introduction that explains the problem statement you are addressing. Why should I be interested in this? **1.2** Provide a short explanation of how you plan to address this problem statement (the data used and the methodology employed). **1.3** Discuss your current proposed approach/analytic technique you think will address (fully or partially) this problem. **1.4** Explain how your analysis will help the reader of your analysis.	5
Packages Required	**2.1** All packages used are loaded upfront so the reader knows which are required to replicate the analysis. **2.2** Messages and warnings resulting from loading the package are suppressed. **2.3** Explanation is provided regarding the purpose of each package.	5
Data Preparation	**3.1** Original source where the data was obtained is cited and, if possible, hyperlinked. **3.2** Source data is thoroughly explained (i.e. what was the original purpose of the data, when was it collected, how many variables did the original have, explain any peculiarities of the source data such as how missing values are recorded, or how data was imputed, etc.). **3.3** Data importing and cleaning steps are explained in the text (tell me why you are doing the data cleaning activities that you perform) and follow a logical process. **3.4** Once your data is clean, show what the final data set looks like. However, do not print off a data frame with 200+ rows; show me the data in the most condensed form possible. **3.5** Provide summary information about the variables of concern in your cleaned data set. Do not just print off a bunch of code chunks with `str()`, `summary()`, etc. Rather, provide me with a consolidated explanation, either with a table that provides summary info for each variable or a nicely written summary paragraph with inline code.	10
Exploratory Data Analysis	**4.1** Uncover new information in the data that is not self-evident (i.e. do not just plot the data as it is; rather, slice and dice the data in different ways, create new variables, or join separate data frames to create new summary information). **4.2** Provide findings in the form of plots and tables. Show me you can display findings in different ways. **4.3** Graph(s) are carefully tuned for desired purpose. One graph illustrates one primary point and is appropriately formatted (plot and axis titles, legend if necessary, scales are appropriate, appropriate geoms used, etc.). **4.4** Table(s) carefully constructed to make it easy to perform important comparisons. Careful styling highlights important features. Size of table is appropriate. **4.5** Insights obtained from the analysis are thoroughly, yet succinctly, explained. Easy to see and understand the interesting findings that you uncovered.	10
Summary	**6.1** Summarize the problem statement you addressed. **6.2** Summarize how you addressed this problem statement (the data used and the methodology employed). **6.3** Summarize the interesting insights that your analysis provided. **6.4** Summarize the implications to the reader of your analysis. **6.5** Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.	5

Section	Standard	Possible Points
Formatting & Other Requirements (These are requirements not a specific *section* written in your report.)	**7.1** Proper coding style is followed and code is well commented (see section regarding [style](#)). **7.2** Coding is systematic - complicated problem broken down into sub-problems that are individually much simpler. Code is efficient, correct, and minimal. Code uses appropriate data structure (list, data frame, vector/matrix/array). Code checks for common errors. **7.3** Achievement, mastery, cleverness, creativity: Tools and techniques from the course are applied very competently and, perhaps, somewhat creatively. Perhaps student has gone beyond what was expected and required, e.g., extraordinary effort, additional tools not addressed by this course, unusually sophisticated application of tools from course. **7.4** .Rmd fully executes without any errors and PDF produced matches the PDF report submitted by student.	15

This grading rubric was modified from http://uc-r.github.io/data_wrangling/final-project