|  |  |
| --- | --- |
| **CSDE 502 2021**  **Assignment 1**  **Introduction to R Markdown**  **Instructor: Phil Hurvitz**  [**phurvitz@uw.edu**](mailto:phurvitz@uw.edu) | **My Name: David Coomes**  **My UWNetID: dcoomes** |

**Due Date: 2021-01-15 09:00**

**Instructions:**

1. Fill out your name and UWNetID at the top of this page.
2. Add your answers to this document.
3. Use the "00Answers" Word style for your answers so they will be clearly discernible from the questions.
4. When you are completed with your work, create a PDF from the Word document.
5. Name your PDF document with the pattern UWNetID\_HW\_n.pdf where UWNetID is your UWNetID and n is the week number when the homework was assigned. For example, the first assignment by me would be named *phurvitz\_HW\_01.pdf*. Upload your completed document to Canvas.
6. Upload any specified data or code files to Canvas.

**Explanation:**

For this assignment, you will be looking in detail at the file week\_01.Rmd (<http://staff.washington.edu/phurvitz/csde502_winter_2021/files/week_01.html>), examining some code that creates a set of files with different names, and suggesting better file names for a list of files.

If you are not familiar with R Markdown format, these resources will be helpful, or others you can find by performing a web search for “R Markdown”:

<https://rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf>

<https://www.dataquest.io/blog/r-markdown-guide-cheatsheet/>

Within RStudio, you can also see:

Help > Cheatsheets > R Markdown Cheat Sheet

Help > Cheatsheets > R Markdown Reference Guide

**Questions:**

1. For week\_01.Rmd:
   1. Explain what the **output** code in the front matter/YAML header of the Rmd file does.

The output code in the front matter/YAML header controls the format of the output files when the R Markdown file is run (knit). In this markdown file there are options for both a pdf and html output, and there are specifications for each output that are detailed in the header including numbered sections, figure captioning, code folding, and several others.

* 1. Explain how the **kableExtra** package is used in the Rmd file.

The kableExtra package is used with the kable package to support additional table formatting. It takes a data frame (irishead in this case) and outputs it as a table with shading and arranges the position on the page.

* 1. Describe how captions and cross-references are implemented in this Rmd with the **captioner** package.

The captioner package references table and figure captions so you can store them as an R code and reference the same figure in your text. It will also automatically renumber your tables and figures, so if you insert a different table above both the text and the table will be updated. This is helpful so that you don’t have to go back through your project and update the numbers every time you make a change.

1. Download and run the file <http://staff.washington.edu/phurvitz/csde502_winter_2021/files/file_naming.R> and briefly comment on different approaches to naming files.

You have to be careful including any spaces in file names. Most of the time it will be okay, but sometimes your program will not run (speaking from experience) and it can be hard to figure out that it’s because of the spaces in your file path. It’s best to not include spaces and just include an underscore whenever you would want one. And it makes sense to put the date as a single number with the year first so that you can order your files by date if necessary.

1. Download the file <http://staff.washington.edu/phurvitz/csde502_winter_2021/files/example_filenames.csv> and produce a table with the original file name and your suggestion for a better file name.

| **date** | **file\_name** | **new\_file\_name** |
| --- | --- | --- |
| 2020-10-13 | Instrument\_v.20 - bilingual.docx | instrument\_bilingual\_20201013.docx |
| 2020-11-30 | Master\_Durham.csv | master\_durham\_20201130.csv |
| 2020-11-30 | Master\_Durham.xlsx | master\_durham\_20201130.xlsx |
| 2020-12-04 | Durham emails\_Nov 30.xlsx | durham\_emails\_20201130.xlsx |
| 2020-12-04 | NJ emails\_Nov 30.xlsx | nj\_emails\_20201130.xlsx |
| 2020-12-07 | Durham tasks\_Nov 30.xlsx | durham\_tasks\_20201130.xlsx |
| 2020-12-07 | Master\_Durham\_v2.csv | master\_durham\_20201207.csv |
| 2020-12-07 | Master\_Durham\_v2.xlsx | master\_durham\_20201207.xlsx |
| 2020-12-07 | NJ emails\_December 7.xlsx | nj\_emails\_20201207.xlsx |
| 2020-12-07 | NJ phone numbers\_Nov 30.xlsx | nj\_phone\_numbers\_20201130.xlsx |
| 2020-12-07 | Pitt phone numbers\_December 7.xlsx | pitt\_phone\_numbers\_20201207.xlsx |
| 2020-12-07 | Pitt phone numbers\_Nov 30.xlsx | pitt\_phone\_numbers\_20201130.xlsx |
| 2020-12-11 | Fraud and quota report\_v1.xlsx | fraud\_quota\_report\_20201211.xlsx |
| 2020-12-11 | Fraud and quota report\_v2.xlsx | fraud\_quota\_report\_v2\_20201211.xlsx |
| 2020-12-11 | Rutgers\_people\_to\_pay\_12\_11\_2020.xlsx | rutgers\_people\_topay\_20201211.xlsx |
| 2020-12-15 | Durham emails\_Dec 14.xlsx | durham\_emails\_20201214.xlsx |
| 2020-12-15 | Durham phone numbers\_Dec 14.xlsx | durham\_phone\_numbers\_20201214.xlsx |
| 2020-12-15 | Export\_1215.7z | export\_20201215.7z |
| 2020-12-15 | NJ emails\_December 14.xlsx | nj\_emails\_20201214.xlsx |
| 2020-12-15 | NJ phone numbers\_December 14.xlsx | nj\_phone\_numbers\_20201214.xlsx |
| 2020-12-15 | VR\_Dec1420\_Durham.xlsx | vr\_durham\_20201214.xlsx |
| 2020-12-15 | VR\_NJ\_Dec1420.xlsx | vr\_nj\_20201214.xlsx |
| 2021-01-05 | transfer\_0105.zip | transfer\_20210105.zip |
| 2021-01-07 | Fraud and quota report\_v3.xlsx | fraud\_quota\_report\_20210107.xlsx |
| 2021-01-07 | Fraud and quota report\_v4.xlsx | fraud\_quota\_report\_v2\_20210107.xlsx |
| 2021-01-08 | test1.dta | test\_20210108.dta |