Building a Code Glossary

## **Description**

The purpose of the code glossary is to take notes and examples in your own words to consolidate the gains you have made in learning how to code in R for data analysis. The code glossary should review and demonstrate mastery of the range of commands, visualization techniques, and analytical methods you have learned this semester. This project is to be completed individually. Use the **Rmd** template provided. You will **hand in the knitted html** document.

You will be graded twice throughout the semester on your Code Glossary progress, but I encourage you to view this assignment as a *living tool* for your own learning and reference throughout the course.

| Due date | What | Submit |
| --- | --- | --- |
| Thursday **Sep 30**  1:30 pm | Initial updates to your code glossary. Provide complete, polished answers to Questions 0-8, 15, 17, 18, 20, 23 (interquartile range and standard deviation will be considered for extra points) | Upload your modified html output to Notebowl, under the “**Code Glossary I**” assignment. |
| Thursday  **Nov 4**  1:30 pm  and  Friday  **Nov 5**  5:00 pm | Updated code glossary. Provide complete, polished answers to Questions 0-4, 10-14, 18, 20-23. You may make revisions to prior answers or examples. This means the only new questions to answer are 0, 4, 11, 14, and 20-23*.* Use code folding and add arguments like include=FALSE to hide code that you don’t need to show. Qs 6, 15, and 19 are optional and may be considered for extra credit (based on the first knitted html turned in before Lab). | Upload your modified html output to Notebowl, under the “**Code Glossary II**” assignment. The lab activity on November 4 will require a completed version of Code Glossary II.  A second deadline will be set for your reflection from the grading activity. |

## **The specifics**

Here, the objective is to **demonstrate mastery** of the skills and tools you’ve developed. This is helpful as a means to review your skills all in one place and be able to use them in a context of your own choosing. Attached is an R Markdown template of the key data wrangling, graphing, and statistical techniques we will use during the semester. For each heading, you will modify the template text and fill in the code chunks to clearly define and produce an example that demonstrates your learning. For example data, you may choose to import one or more example datasets, or use a dataset from one of the many R data packages that are freely available (e.g. titanic dataset)\*. You can use one dataset for all your examples to simulate a data analysis that really digs into one topic, or you can switch it up if you like variety.

**Data:**

Some examples of datasets you could use are below. **Take a few minutes to explore and choose something that interests you!** Making your code glossary will be more motivating if you’re actually interested in the data examples you choose.

* Free sources of data online: <https://r-dir.com/reference/datasets.html>
* Datasets that we’ve used in a lab or in class\*
* Data from one of the R data packages, like the NHANES data we’ve used in class
  + storms (tidyverse)
  + starwars (dplyr)
  + Run data() in your console to see what datasets might be pre-loaded

\*If you choose to work with a dataset that we have already been introduced to in a lab, **command lines that were given in previous labs or in the class notes may not be directly copied**---you must apply the command to a different variable in the dataset or use the variable with a new command, or modified in some way. Similarly, **examples may not be taken verbatim from the web**. Doing so does not demonstrate mastery and understanding, and it is plagiarism.

*You should hide all setup or intermediate code that is not part of the specific functions you are demonstrating. In other words, it should be clear to your professor which parts are being demonstrated for the assignment. Dr. Supp can show you how to modify code chunks to hide output, as needed.*

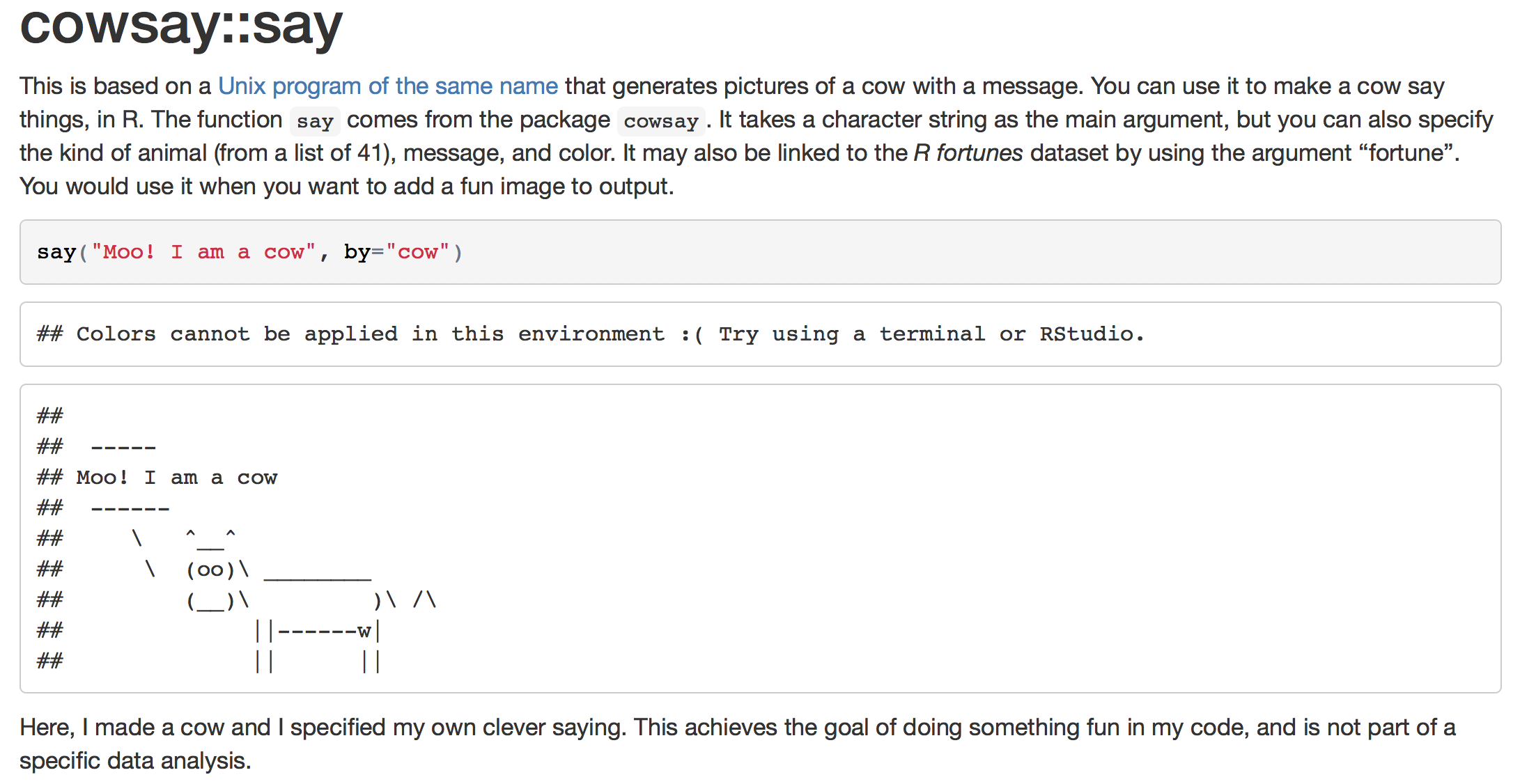
**For each code chunk follow the template provided:**

1. Name the function(s) you are demonstrating in the code chunk.
2. **Package**: Name what library the function belongs to.
3. **Definition**: In 2-3 sentences, define the function that you are demonstrating. This should be a general definition, that would tell a new R user what the function would be used for, when it would be appropriate to use, and/or what the technique accomplishes. Imagine you are explaining the function to a friend who has never taken Data Analytics.
4. **Code example**: Your example code demonstrating this function using a dataset -- provide evidence that your code worked. If you are using code folding, this may be present under the “button” in the knitted document. If you are not using code folding, make sure to show the code. For all graphs and tables, annotate and polish (make them presentation-ready). The code should follow basic R style guidelines and best practices.
5. **Explanation**: 2-3 sentences of interpretation, specific to the example and the data that you just showed. *What did the code accomplish in the example that could be useful in a data analysis?* Imagine that you are explaining the code, function, and/or plot to a friend that has never taken Data Analytics.

I recommend that you work on your code glossary throughout the semester, not just before one of the graded check points. Building your code glossary throughout the semester is an excellent way to solidify and practice your knowledge, and have a living guide to the code that you can reference for help later. While all students will be building a code glossary of the same functions, no two students should have identical examples, definitions, or explanations. Please avoid the temptation to share the code directly with each other, or to workshop your individual text and answers together, unless specifically instructed to do so.

**You must work on your own and explanations must be in your own words. Do not copy and paste text from an online (or any) reference. Zero credit will be given for copied definitions or examples. Commands and/or interpretations that are identical among class members, or that are copied verbatim (or close to verbatim) from another source will be reported as a breach of Denison’s academic integrity policy.**

An example is below, though you should use as much space or detail as you need to clearly describe and define the function.

****