Data Science: Fundations using R

What is Data Science: Cuestionario práctico

1. Which of the following is an example of structured data?
   A database of individual's addresses, phone numbers and post codes.
2. Which is **NOT** one of the three V's of Big Data?
   Vast, versatile, vital.
3. Which of these is **NOT** one of the main skills embodied by data scientists?
   Access to large datasets.

What is Data: Cuestionario práctico

1. Which of these is an example of a quantitative variable?
Weight, Height, Color.
2. Quantitative variables are measured on ordered, continuous scales. True.
3. What is the most important thing in Data Science? The question you're trying to answer.

Getting Help: Quiz

1. Which of these might be a good title for a forum post?
   Removing rows with NAs in data.frame using subset(), R 3.4.3.

2. Which is a characteristic of a good question when posting to message boards?
   Is polite and corteous.

3. Which is **NOT** a good strategy for finding the answer to a problem in this course?
   Emailing the professor.

Data Science Process: Quiz

1. Which of these is **NOT** an effective way to communicate the findings of your analysis?
Save code locally on your computer.

2. What's the first step in the data science process?
   Generating the question.

3. Why should you include links or citations to others' work?
   It helps other quickly find information about you've reference.

4. What does Hilary Parker suggest led to the popularity of the name "Dewey" in the late 1800s? (You may have to reference Hilary's blog post for this question.
   People named her daughters after George Dewey, after the Spanish American war.

Installing R: Cuestionario Práctico

1. What does CRAN stand for?
   Comprehensive R Archive Network.
2. What does base R focus on?
   Statistical Analysis.
3. What is the output when you type into R: mean(mtcars$mpg) ?
   20.09062.
4. Why are we using R for the course track?
   R is one of the most widely used programming languages for data science.

Installing R Studio: Cuestionario

1. What is RStudio?

A graphical user interface for R.

2. Which is **NOT** an option for a file type when you go to File > New File in RStudio?

R Beamer presentation.

Installing R Studio: Cuestionario

1. How do you see a command you have previously run and save it to source?
   History tab > Highlight command > To source.
2. What is the name of the quadrant in the bottom left corner of RStudio, in the default layout? Console
3. Which of the following is **NOT** one of the options available under the Global Options menu in Tools? Versions.
4. Using the Help menu, find out which of the following is one of the three species of Iris present in the base R dataset: iris. Virginica

R Packages: Cuestionario

1. How would you install the package ggplot2? Install.packages("ggplot2")
2. Using the help files, what is **NOT** a function included in the devtools package? aes().
3. Which is **NOT** one of the main repositories? RDocumentation.
4. What command lists your R version, operating system, and loaded packages? sessionInfo()
5. Install and load the KernSmooth R package. What does the copyright message say?

```
Copyright M. P. Wand 1997-2009
```

Projects in R: Cuestionario

1. Which is **NOT** a way to create a new Project? Session > New Project
2. What file extension do Projects in R use? .Rproj
3. Creating a new project from scratch will **NOT** do which of the following? Initiate version control.

Projects in R: Cuestionario

1. I'm done editing a file, I need to _____ those changes then _____ them, and _____ it to the _____.
   Stage, commit, push, repository.
2. What is a good example of a message to accompany a commit?
   Modified lineal model of height to include new covariate, genotype.
3. Which of these is **NOT** true about using a version control system?
   Version control is time-consuming and difficult to use.

Github and Git: Cuestionario

1. On each repository page in GitHub, in the top right hand corner there are three options. They are:
   Watch, star, fork.
2. To make a new repository on GitHub, which can you **NOT** do?
   Profile > New repository.
3. What command can you use to change the name associated with each of your commits?
   Git config –global user.name “name”
4. What command can you use to see your Git configuration?
   Git config –list
5. Which of the following will initiate a git repository locally?
   Git init

Linking Github and RStudio: Cuestionario

1. In what quadrant of RStudio will you find the Git tab? Environment.
2. What is the order of commands to send a file to GitHub from within RStudio?
   Stage > Commit message > Commit > Push.
3. Which can you **NOT** do from within the Commit window of RStudio?
   None of the above.

Projects under versiôn control: Cuestionario

1. What do you call it when you create a local copy of a repository that you will work on collaboratively with the original repository owner? Clone.
2. What is the command to initialize git in a directory? git init.
3. How do you make a commit from within the command line? git commit -m “message”

R Markdown: Cuestionario práctico

1. How would you strike through some text? ~~strikethrough~~
2. What is the format for including a link that appears as blue text in your markdown document?
   [text that is shown](link.com)
3. How do you produce **bold** text? **bold**
4. How do you produce *italicized* text? *some text*
5. How do you produce your final document? Knit.

Types of data Science questions: Cuestionario

1. Which of the following describes a predictive analysis?
   Using a data collected in the past to predict values in the future.
2. We collect data on all the songs in the Spotify catalogue and want to summarize how many are country western, hip-hop, classic rock, or other. What type of analysis is this? Descriptive.
3. We collect data on a small sample of songs from the Spotify catalogue and want to figure out the relationship between the use of the word "truck" and whether a song is country western. What type of analysis is this? Inferential.

Experimental Design: Cuestionario:

1. In a study measuring the effect of diet on BMI, cholesterol, lipid levels, triglyceride levels, and glycemic index, which is an independent variable? Diet.
2. Which of the following is **NOT** a method to control your experiments? Placebo effects.
3. What might a confounder be in an experiment looking at the relationship between the prevalence of white hair in a population and wrinkles? Age.
4. According to Leek group recommendations, what data do you need to share with a collaborating statistician? All the above.
5. If you set your signifance level at p-value ≤ 0.01, how many significant tests would you expect to see by chance if you carry out 1000 tests? 10.
6. What is an experimental design tool that can be used to address variables that may be confounders at the design phase of an experiment? Stratyfying variables.
7. Which of the following describes a descriptive analysis?
   Generate a table summarizing the number of observations in your dataset as well as the central tendencies and variables of each variable.

Big Data: Cuestionario:

1. Which is **NOT** one of the three V's of Big Data? Valuable.
2. Which one of the following is an example of structured data? A table of names and student grades.
3. What is the reason behind the explosion of interest in big data?
   The price and difficulty of collecting and storing data has dramatically dropped