Scout (Mugdha) Sujyot

Week 3 paper

What are the top 3 ideas you retained from the material read in chapter 2?

1. When importing data in R, it's highly recommended to write scripts and automate steps. Often data isn't well-structured and will need transformation for it to make sense.
2. Data comes in many kinds of files, but it often lives in relational databases. You often have to curate this data and it's very important to keep notes, scripts, and code as you must be able to defend your results and repeat your work
3. Don't be afraid to work with samples of your data instead of all of it. Often your results will be the same with a reasonably sized data sample and you can save much effort in working with a sample.

What are the top 3 ideas you retained from the material read in chapter 3?

1. You can use the summary command to explore your data and spot potential problems, like missing data or unlikely values. It also gives you a rough overview of your data.
2. You can also use visualisation to explore your data. In doing so, you should try to display as much information you can as clearly as possible.
3. There are many different graph types you can use and you should choose them based on the kind of relationships you're looking at. For instance, if you're looking at the relationship between two continuous variables you might want to use a line plot, a scatter plot, a smoothing cure, or a hexbin plot. However, if you're looking at two categorical variables, you may want a stacked, side-by-side, or filled bar chart. The specific type of relationship is important to consider too. If your relationship is functional or nearly so, a line plot is a good bet -- however, if the relationship is too loose or cloud-like to be seen on a line plot, a scatter plot might be a better option.