Riley Payung

CDS 301

04/20/2020

Quiz 4

Q1

The video is a demonstration of the creator's package to bootstrap dataframes in R onto ggplots. It allows you to linearly regress data to find trends based on specified types, such as the type of iris. It also allows you to animate using gganimate much more nicely. The data can also be normalized. Whole point is to be able to model uncertainty with as much accuracy as possible.

Q2

The article discusses the pros and cons of individual visualization types, mainly studying what is more accurate at 1. Modeling uncertainty, and 2. Effectively displaying the information in the easiest and fastest way possible. Intervals are good at allowing a reader to visualize quickly, but are unable to model uncertainty; they are exact, and are ambiguous in nature. Probability density maps are good at providing a quick look at a distribution of data, making it look more normalized while using another type of visualization such as a scatter graph or line graph, whereas they tend to show bias in some spaces, namely causing the reader to estimate the data incorrectly, using the darkest or widest areas. The article then goes on to mention the hybrid approaches of data visualization, combining some visualizations to make the data easier to visualize; however, this has its own cons because there is uncertainty in the visualization that the reader will pay the most attention to, even though there could be many visualizations in the entire paper or dashboard.

Q3

The video shares the AFFECT of using color in visualizations based on the audience you're presenting to, where a business visualization might be more muted than in a market with more liberal arts, such as the game industry, which would tend to be more vibrant (They like colors, trust me). The idea is to make the audience feel a certain way or impress. Colors need to be distinct for building visualizations because too many colors make it more difficult to understand categorical data.