R Project

Web Application

DMDS 2019-2020 - Tania QUEUCHE

https://tania-queuche.shinyapps.io/OBUS/

Process

The project was completed along with R class from September to December 2019.

1. **Finding a database** (passive process, approx. 1 month looking on and off).
2. **Exploring dataset and plotting graphs, thinking about the layout and storytelling** (October-December). I decided to start with the graphs (instead of the appli) as I wanted to see what could be analyzed and plotted, then design the application in function of that.
3. Coding the **web application** (28/11-15/12)

Layout (storytelling, tabs)

The shiny application contains 4 sections contained in a shiny app Tab: *General, Correlation, Behavior* and *Socio-economic*. In each section, I analyze a different aspect / factor of obesity. The sections should be explored in order as together, they constitute the storytelling.

Section	Description	Main functions used
General	Introduction: general overview on obesity in the US. Facts from the database (obesity rates by state, mean obesity rate in the US) and facts retrieved from external sources (in the Information Boxes). Hopefully this first tab will surprise the user and make them want to learn more.	*Infobox, geom_col, slider (number of states shown), shinydashboardPlus*
Correlation	This is an informational tab – there's no need for a lot of interactive content. The aim is to show the correlation between 3 behavioral factors and their correlation with obesity. The choice of graphic was quite easy: for a correlation it seemed obvious to use a geom_point, and the log_scale function enabled to see the correlation more easily.	*Geom_point (logarithmic scale), dataframe*

Behavior	Now that the user knows 1) the obesity rates in each state 2) the causes leading to obesity, the user is invited to explore the behavior by state. It may be confusing to interpret the graph at first because it represents the % people NOT practicing any physical activity regularly. The higher the bar is, the more people do NOT practice sports in a given state, the laziest it is. Same for consumption of fruits and vegetables (the higher the bar, the less people consume greens daily). To make it more intuitive, I customized the colors: - For sports, the laziest states are in red as the color has a negative connotation. - States where people don't eat much greens are in yellow-ish, and gradually becomes greener	*Geom_col, manual color fill, interactive display (number of states shown), HTML tags*
Socio-eco	In this section, the user is introduced to a new sort of factor: socio-economic ones. The graphs show the relation between Income / Education level / Gender and obesity. Due to the scale, the exact percentages (obesity rates) are hard to read, so the graphs were made in ggplotly with a legend when hovering.	*Ggplotly, manual color fill, pivot table, HTML tags*

Header	Displays links to my LinkedIn page and to the database. Notification icon (not very useful but just to show I can do it)	*Hyperlink tags (tags$li), dropdownmenu, icon, messageitem*
Sidebar	Contains the "chapters" of the storytelling and the interactive slider. The slider works on any histogram featuring states.	*Menuitem, slider*

Final thoughts

- To be **resilient**. It sometimes takes hours to solve one single problem (from finding what the problem is, then skimming through a multitude of articles/forums/videos on Google to find which ones are relevant, then finding out how to apply it to my own code).
- To seek for **help**. It is useful to have a fresh look on what one is doing. Others can see things we don't when we've been looking at the same code for hours. And sometimes the errors are silly (like a typo).
- To be **realistic**. When looking at tutorials it is tempting to do impressive things, but sometimes it is not realistic to do so (because my level in R is not advanced enough or because it would take an unreasonable amount of time to do it)
- If I had to do this again, I would have named the variables better (with a consistent logic – it was difficult to remember the names of all my variables because there was weird names)
- **Data analysis is certainly not easy, but it is rewarding**. Looking back I realize I have progressed a lot.