W12-submission

Janelle Tan

2023-11-08

W12 submission

A.Diary Entry

1. What is the topic that you have finalised? (Answer in 1 or 2 sentences).

```
# I have decided on the topic of "health" for my data story, namely, Alcohol Abuse and its Publi
c Health Consequences.
```

2. What are the data sources that you have curated so far? (Answer 1 or 2 sentences).

```
# My data sources for now are mainly coming from the Global Health Observatory data repository.

# Some data sets I'm considering are regional prevalence of alcohol dependence in 2016 (%), Regi
onal prevalence, AAFs (15+), liver cirrhosis deaths (%), Alcohol, recorded per capita (15+) cons
umption (in litres of pure alcohol).
```

3. What is the question that you are going to answer? (Answer: One sentence that ends with a question mark that could act like the title of your data story)

```
# "What is the extent of alcohol abuse and its public health consequences in the Americas?" (Not
e: question is still subject to change)
```

4. Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., “According to the United Nations…” to justify why the question you have chosen is important)

```
# Substance abuse is a significant public health concern with severe consequences. According to the United Nations, substance abuse contributes to over 350,000 preventable deaths annually. Moreover, the World Health Organization (WHO) notes that substance abuse leads to a significant economic burden, with costs related to healthcare, law enforcement, and lost productivity. This question is important as understanding the global impact of substance abuse can inform effective public health policies and interventions.
```

5. Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use). Include the challenges and errors that you faced and how you overcame them.

```
# Note: the below are still subject to change

# Data Sets: Regional prevalence of alcohol dependence & liver cirrhosis, Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)


# Specific Variables for Regional prevalence of alcohol dependence:
# Columns: WHO region, regional prevalence of alcohol dependence
# Rows: all

# Specific Variables for Regional prevalence of liver cirrhosis:
# Columns: WHO region, regional prevalence of liver cirrhosis
# Rows: all

# Specific Variables for Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol):
# Columns: Regions, BeverageType, Year 2018
# Rows: all


# Challenges and Errors:
# So far, I've encountered some challenges trying to narrow down the scope of my question. It was also difficult trying to find suitable data sets for the scope of my question amidst the sea of data available.

# To solve them, I narrowed down the scope of my question based on what data was available. I learnt how to quickly filter out potential data sources by looking at its size and whether or not its variables were relevant to my question.
```

6. List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

```
# Bar graphs.

#Variables include WHO regions, regional prevalence of alcohol dependence (%) and liver cirrhosis (%) both in 2016.
#For the data set on Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol), variables include: Regions, BeverageType, Year 2018

# The bar graphs show how the alcohol scene and consequences of alcohol abuse in the Americas are the one of the worst among the other regions, highlighting the extent of substance abuse there.
```

7. How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

```
# I'm currently experimenting with plotly and ggiraph packages to make my bar graphs interactive. I'm also contemplating using shiny apps.
```

8. What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

```
library(tibble)

# Create a tibble with columns for topics, weeks, and self-learned
course_concepts <- tibble(
  Topic = c("Visualising data using ggplot", "Making interactive graphs: plotly & ggiraph"),
  Week = c(7, NA),
)

# Print the tibble
print(course_concepts)
```

```
## # A tibble: 2 × 2
##   Topic                                        Week
##   <chr>                                       <dbl>
## 1 Visualising data using ggplot                   7
## 2 Making interactive graphs: plotly & ggiraph    NA
```

9. Challenges and errors that you faced and how you overcame them.

```
# I had trouble making my bar graph interactive with plotly.

# I overcame them by referring to websites like stack overflow for solutions and trying out diff
erent ways to make my plots interactive, like with ggiraph and shiny apps. I am currently still
experimenting with these methods and looking into other possible methods.
```

10. Challenges and errors that you faced and how you overcame them (Week 12).

```
# I was looking over my question again and realised that I wanted to make my question and answer
more meaningful if possible. I started looking for ways to do so, though along the way, I strugg
led with coming up with suitable ideas.

#I overcame this challenge by looking at available data sets to see if there was anything I coul
d use to achieve my goal. I am currently still on the look out for more ideas, though I have som
e rough plans, including adding in data sets that could help me further assess the impacts of al
cohol abuse (eg looking at the different types of health consequences that result from alcohol a
buse).
```