Final Report

Molly Hischke

12/9/2019

# Introduction: Research question, and Why did you decide to ask this question?

# Methods:

How did you investigate the data to try to answer your question? This should not include R code (save that for the tutorial part), but rather should use language like “To determine if … was associated with …, we measured the correlation …”. It’s fine for this project if the Methods are fairly simple (“We investigated the distribution of … using boxplots …”, “We took the mean and interquartile range of …”, “We mapped state-level averages of …”, etc.). Why do you choose to use the Methods you used? Why do you think they’re appropriate and useful for your project?

## -- Attaching packages -------------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.2  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ----------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

There were two main challenges in finding the positions of the daily double questions:

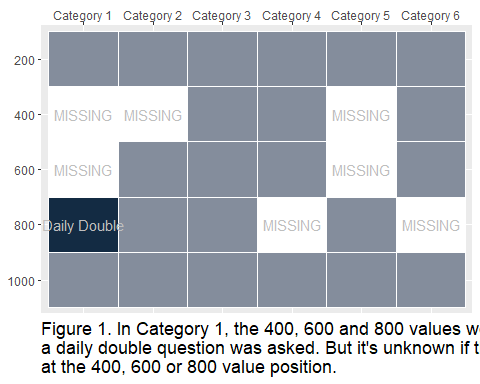
1. Finding the x position (the location of the categoires).
2. Finding the y position (the location of the values).

Not all categories per episode (n = 12) were available in the dataset. To solve this issue, only episodes where all 12 categories were played were included. The assumption was that the dataset listed the categories in order by location per episode (i.e. Category 1, Category 2, etc.).

In addition, not every question (n = 5) within a category was availabe in the dataset. This was espeically challenging when a category contained a daily double question. Since players select their own bid for daily doubles, the value within the dataset gave little information about the location on the y axis. If there were 2 missing questions and 1 daily double within a cateogyr, the daily double could be at any one of those locations (Figure 1).

jeopardy\_board %>%   
 mutate(y\_pos = factor(y\_pos),   
 y\_pos = factor(y\_pos, levels = rev(levels(y\_pos)))) %>%   
 ggplot(aes(x = x\_pos, y = y\_pos, fill = number\_of\_dd)) +  
 geom\_tile(color = "White", size = 0.1) +  
 scale\_fill\_continuous(high = "#132B43", low = "white") +   
 labs(x = NULL,   
 y = NULL,   
 fill = NULL) +  
 theme(legend.position = "none") +   
 geom\_text(aes(label = text), color = "gray") +  
 scale\_x\_discrete(position = "top") +  
 labs(caption = paste0("Figure 1. In Category 1, the 400, 600 and 800 values",  
 " were missing, and \na daily double question was asked. But it's unknown if ",  
 "the daily double was \nat the 400, 600 or 800 value position.")) +  
 theme(plot.caption = element\_text(hjust = 0, size = rel(1.2)))

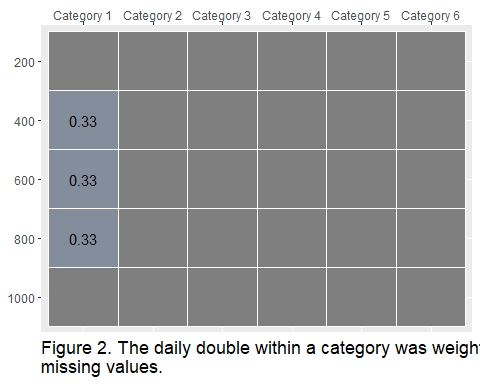
## Warning: Removed 22 rows containing missing values (geom\_text).



To solve the issue, the daily double was looked at as a ‘missing value’, and the y position of the daily double was weighted across the three ‘missing’ locations (Figure 2).

weighted\_jeopardy\_board %>%   
 mutate(y\_pos = factor(y\_pos),   
 y\_pos = factor(y\_pos, levels = rev(levels(y\_pos)))) %>%   
 ggplot(aes(x = x\_pos, y = y\_pos, fill = number\_of\_dd)) +  
 geom\_tile(color = "White", size = 0.1) +  
 scale\_fill\_continuous(high = "#132B43", low = "white") +   
 labs(x = NULL,   
 y = NULL,   
 fill = NULL) +  
 theme(legend.position = "none") +   
 geom\_text(aes(label = number\_of\_dd)) +  
 scale\_x\_discrete(position = "top") +  
 labs(caption = paste0("Figure 2. The daily double within a category was ",  
 "weighted acrossed all \nmissing values.")) +  
 theme(plot.caption = element\_text(hjust = 0, size = rel(1.2)))

## Warning: Removed 27 rows containing missing values (geom\_text).



# Results:

What did you find out? Most of these slides should be figures or tables. Discuss your interpretation of your results as you present them. Ideally, you should be able to show your main results in about 3 slides, with one figure or table per slide.

# Conclusions:

So what? How do your results compare with what other people have found out about your research question? Based on what you found, are there now other things you want to check out?