Lab 2 Activity

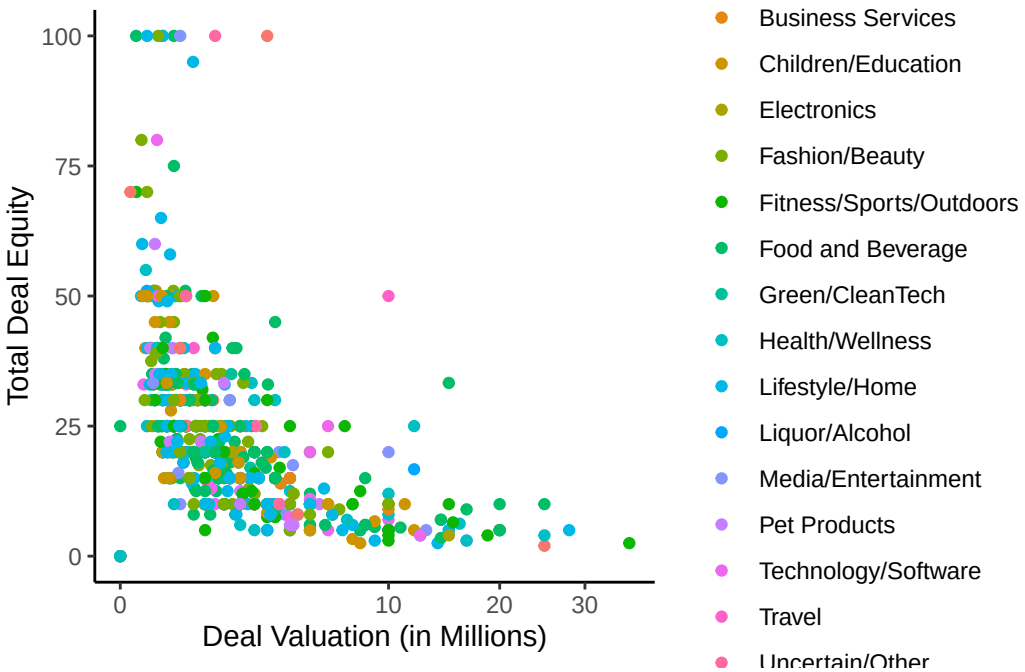
Use Quarto to create a pdf document of your results. When you are done (or when lab time runs out), submit both your code document and pdf document to Blackboard. **This assignment is not graded aside from participation points.**

First make sure that you download the `Shark_Tank_US.csv` dataset. This is data about the US version of Shark Tank, a business reality TV show. Make sure that this file is in your working directory (you may want/need to change your working directory).

```
library(tidyverse)
ST <- rio::import("Shark_Tank_US.csv")
```

Recreate the following graphs to the best of your ability (feel free to start with any one you please). Use the `ggplot` help pages and the internet to help you! We don't expect you reproduce every aspect of every plot today, but we want to give you some real-world problem-solving practice. After lab today, we'll give you the source code for you to review on your own time.

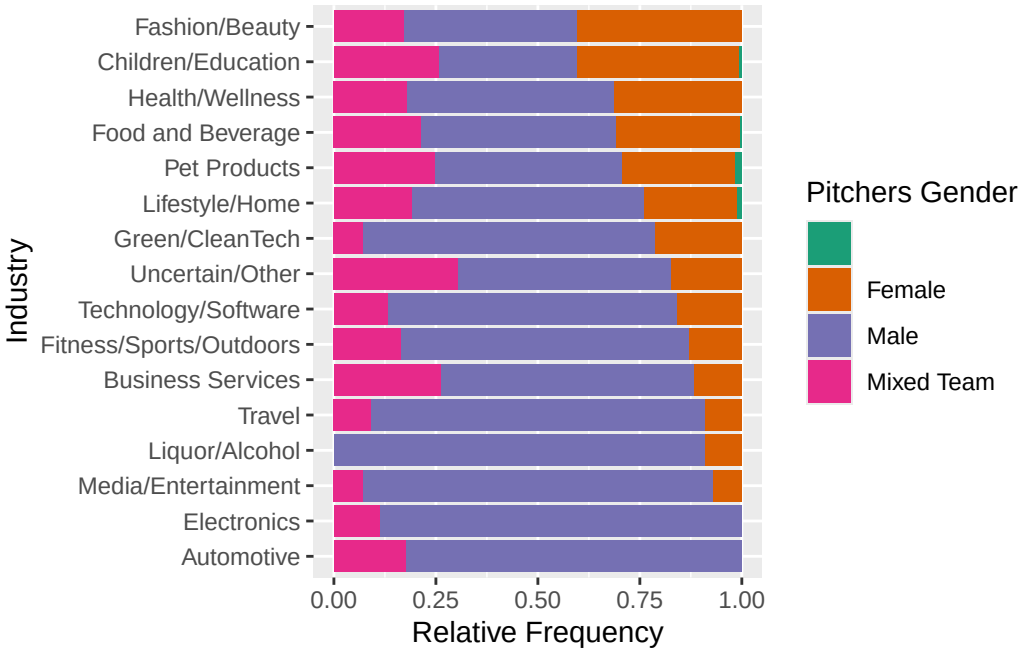
Plot 1



Plot 1 checklist:

- Filter only those pitches that received deals
- Transform Deal Valuation (in USD) to millions of USD
- Plot different industries in different colors
- Create a scatterplot with deal valuation on x-axis and total deal equity on y-axis
- Make sure x-axis label matches the plot
- Choose a plotting theme that omits the background grid lines
- "Stretch" the x-axis scale using square root scaling

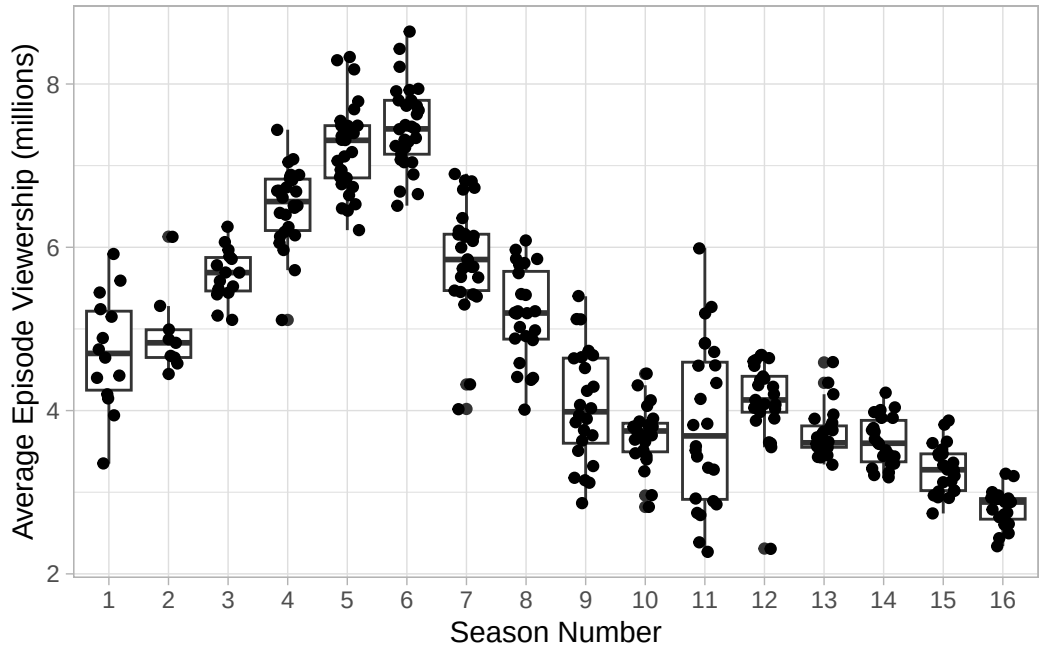
Plot 2



Plot 2 checklist:

- For each industry, calculate the proportion of pitchers that are female. Use the argument `na.rm = TRUE` to avoid `NA` proportions when the average involves `NA` values of gender
- Reorder the levels of the `Insdustry` variable by levels of Proportion Female (previous step)
- Create a stacked bar plot that fills the plotting space
- Use different colors for different genders
- Different industries should be in different rows
- Match the x and y-axis labels
- To match colors exactly, use the "Dark2" palette from the `RColorBrewer` package (see `scale_fill_brewer` function in `ggplot2`)

Plot 3



Plot 3 checklist

- This plot includes one point per **Episode** rather than one point per **pitch**. Therefore, before plotting we want to filter the data so that there is only one row per episode. There are several ways to do this.
- One episode has missing values of `US Viewership`. Omit this episode from the plot.
- Season number should be converted to a factor type variable
- Create boxplots
- Create jittered scatterplots with width of .2
- Choose a theme that matches the picture
- Match x and y-axis labels