# MATH 158 - Data Description and Descriptive Statistics

John King

due Thursday, Feb 8, 2022

### Introduction to Data

The data from this project comes from TidyTuesday and fivethirtyeight on GitHub. This data is a collection of the top 10 companies who posted the most advertisements during the Superbowl between the years 2000 and 2020. Because of errors in data collection, seventeen of the videos had to be removed from the data. From this, we want to compare the different qualities in the video to the social media performance after the event.

The observational unit is each individual advertisement, and there are 10 variables. The categorical variables are logical binary variables which identify as True or False.

Quantitative Data

- Year: This variable indicates the year the Superbowl advertisement aired on TV.
- View Count: This variable indicates how many youtube views the advertisement has received.
- Like Count: This variable indicates how many youtube likes the advertisement has received.
- Dislike Count: This variable indicates how many youtube dislikes the advertisement has received.
- Comment Count: This variable indicates how many youtube comments the advertisement has received.

Categorical Data

- Funny: This variable indicates whether the advertisement is intended to be funny.
- Celebrity: This variable indicates whether a celebrity is in the advertisement.
- Danger: This variable indicates whether there is danger in the advertisement.
- Animals: This variable indicates whether there are animals in the advertisement.
- Use Sex: This variable indicates whether there is use of sexuality in the advertisement.

### Summary of Statistics

As mentioned above, the dataset looked at the 10 companies who prepared advertisements between 2000 and 2020. Out of these 10 companies, this is how many total advertisements they posted each year.

```
maindata <- read.csv("Data - ExportedData.csv")
table(maindata$year)
```

```
##
## 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
##    6   12    9   10   11    6    7   14   13   15   12    9   15   11   11    8
## 2016 2017 2018 2019 2020
##    9    5   12   12    9
```

Next, I want to reveal the number of observations which identify as True or False for every categorical variable.

| | Funny | Celebrity | Danger | Animals | Use_Sex |
|---|---|---|---|---|---|
| True | 171 | 71 | 75 | 92 | 66 |
| False | 45 | 145 | 141 | 124 | 150 |

```
maindata <- read.csv("Data - ExportedData.csv")
```