

# STA325 Case Study

Chase Mathis

## Packages/Data

```
library(tidyverse) # datawrangling
library(tidymodels) # modeling steps
library(knitr) # nice output

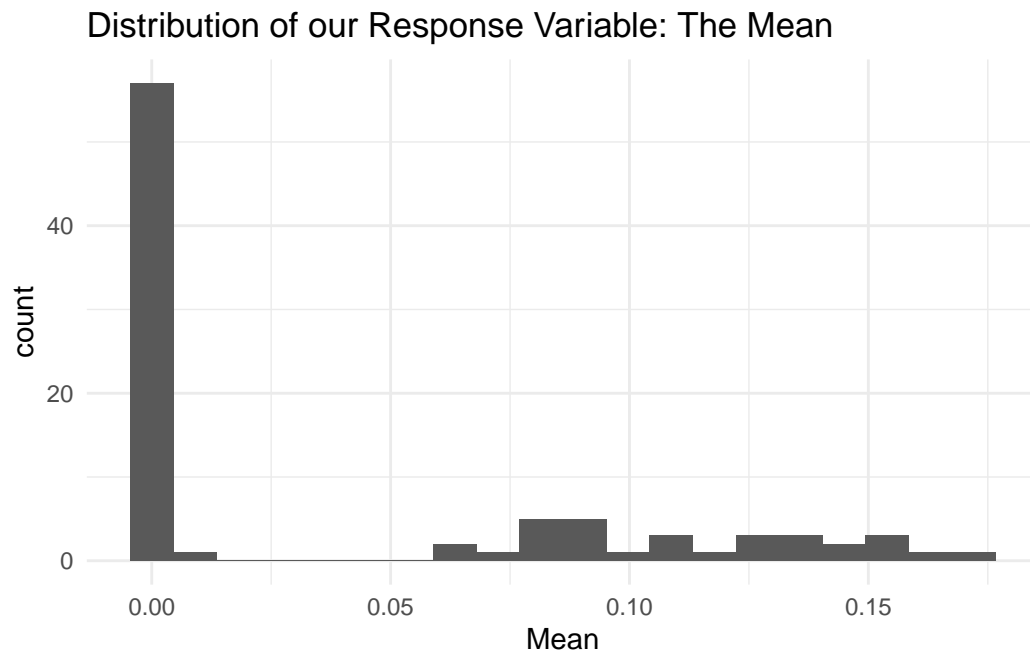
ggplot2::theme_set(ggplot2::theme_minimal())

train <- read_csv("data/data-train.csv")
test <- read_csv("data/data-test.csv")
```

## Model for the Mean

As there are multiple response variables, I believe making individual models for each one will work best for both prediction and inference. We start with some EDA.

## Exploratory Data Analysis

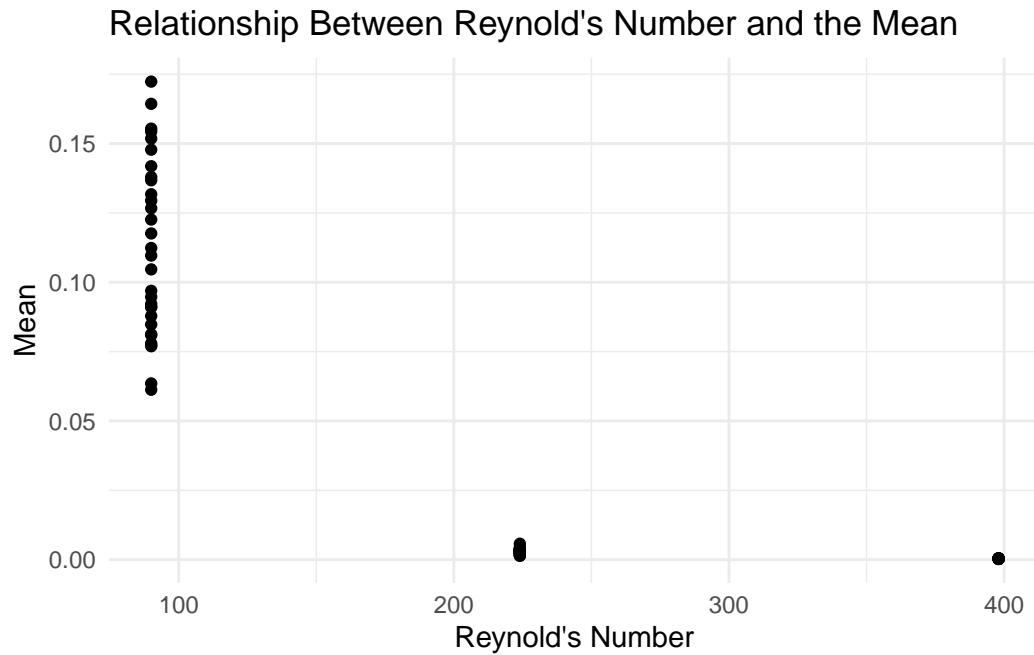


The majority of the first raw moments, i.e. the mean, are very close to zero. This may impact our model, and is something to look out for.

## Investigating Relationship Between Reynold's and Mean

Table 1: Re Distribution

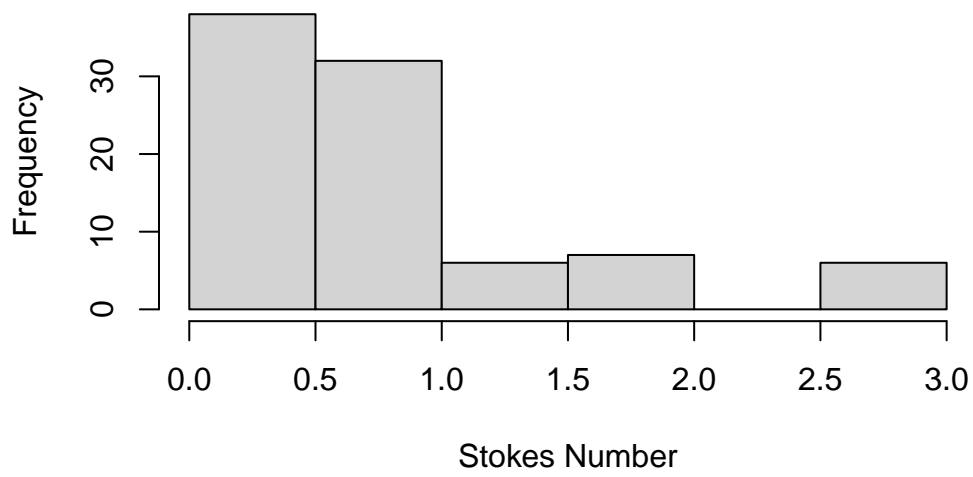
Re	Count
90	31
224	39
398	19



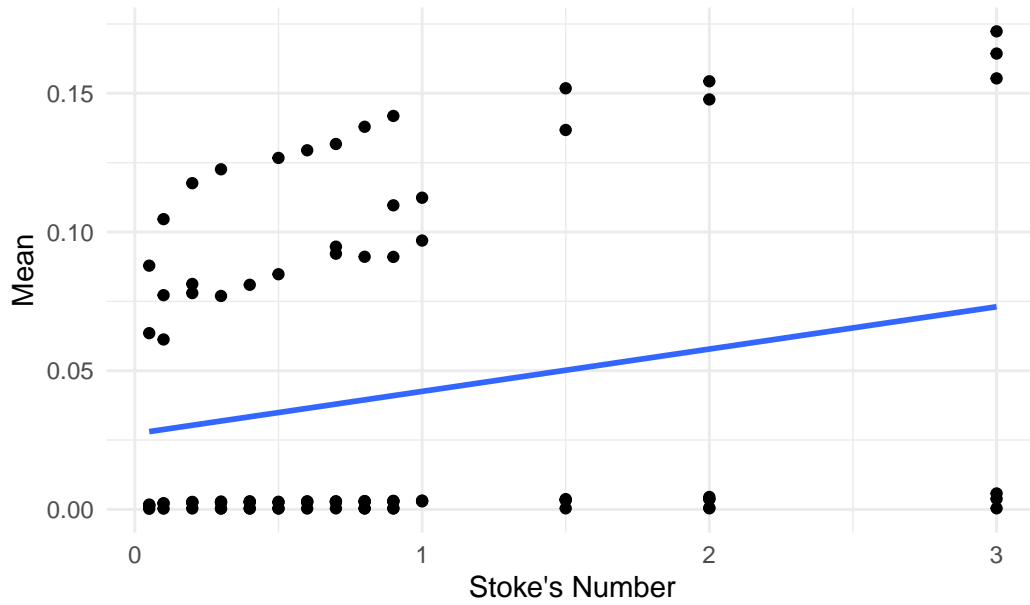
There are three different values for Reynolds number as seen in the table. Then, when looking at the visualization, it looks like if Reynolds number is the two larger numbers, the mean is very close to zero.

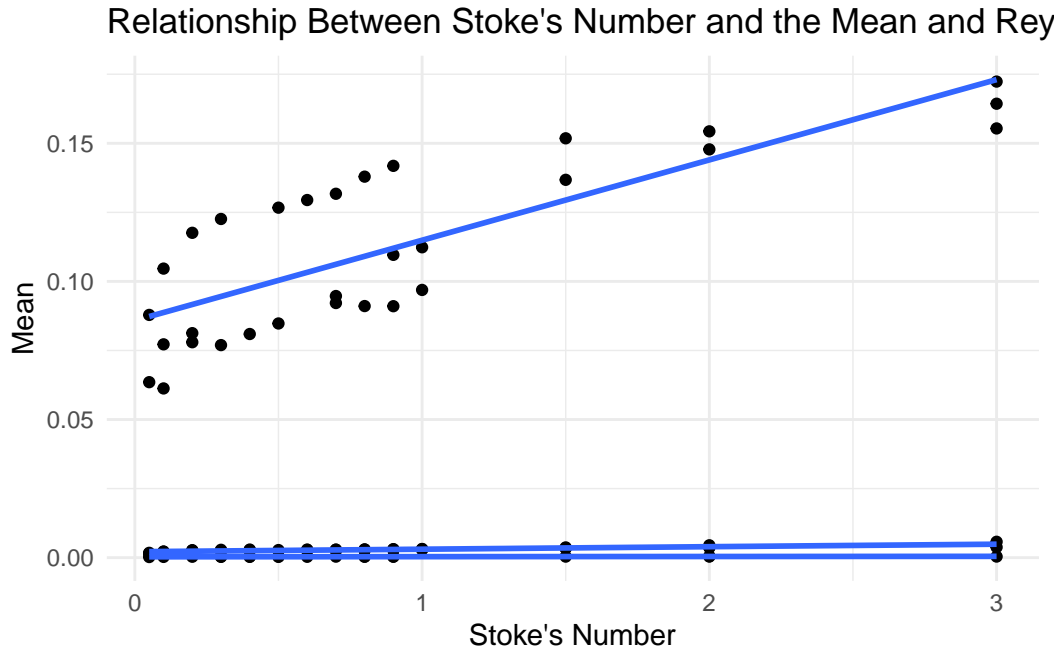
### Investigating Relationship Between Stokes Number and Mean

**Histogram of Stokes Number**



**Relationship Between ONLY Stoke's Number and the Mean**





If we fit a linear model with only Stoke's Number, it looks pretty terrible, but if we add in Reynolds number it becomes much better. We noticed a small trend in the plot above, but adding Stokes number takes away some uncertainty and explains when Reynolds number is small.

### Investigating Relationship Between Mean and Froude's Number

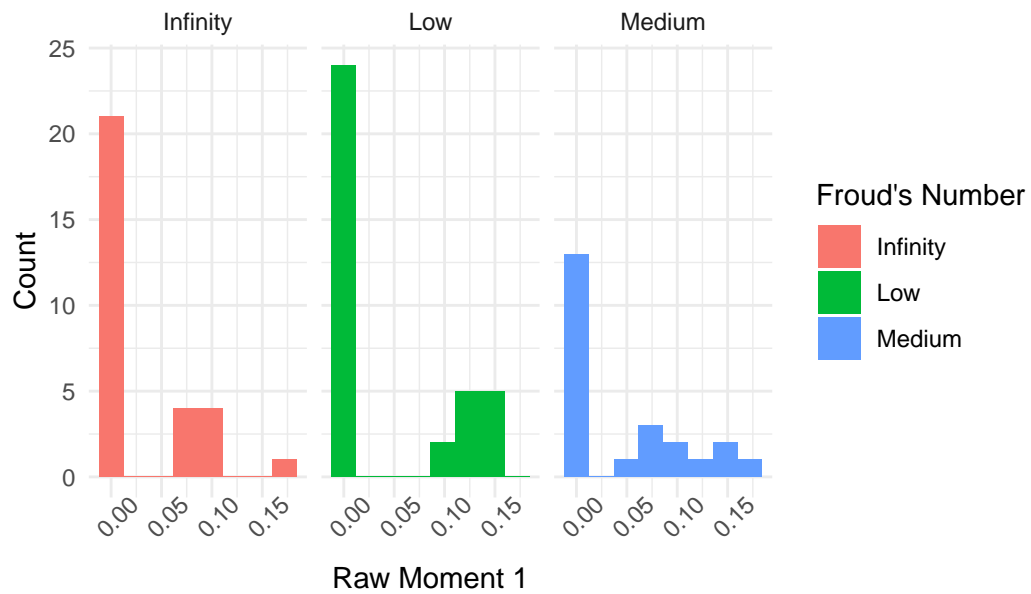
Table 2: Distribution of Froude's Number

Froude's Number	Count
0.052	36
0.300	23
Inf	30

Similar to Reynolds number, this number is somewhat non-continuous as there is only three buckets for the number. Therefore I am going to make it a categorical variable to help interpretability. For instance, how do we interpret Infinity? Instead I will make

- 0.052 == Low
- 0.300 == Medium
- Inf == Infinity

## Relationship between Froud's Number and the Mean



It seems hard to distinguish what affect Froud's number has on the mean from this plot.