Faith Adade/Josh Wufsus

Checkpoint 2

1. Guiding Questions

- How is coral reef health changing over time?
- Do the changes in ocean acidity and coral calcification positively correlate?
- Which of the ten sampled coral reefs have experienced the greatest changes in growth (calc, density, extension)?
- What geographical regions are being represented in the data collection?

2. Introduction

Our work examines the changes in the Great Barrier Reef over the span over 50 years. In addition, we are looking at water composition and major pollutants that have likely affected the health of coral reefs. By creating visualizations and averaging our data, we will be able identify which variables have strong correlations with the damages of our major coral reef systems.

3. Primary Dataset

1. The primary data set was sourced from the NOAA (National Oceanic and Atmospheric Administration). The organization obtained this data by collecting coral cores and implementing paleoclimatology techniques to gather information. The data collected ranges from 1810-2000, depending on the coral sample.
2. The data reports accurate data from a wide span of time, which allows us to observe long term trends in coral health.
3.

```
t <- read.csv("Desktop/project/stoneh.csv")
head(t)
```

```
  Year   Ext   Den  Calc
1 1955 1.727 1.278 2.207
2 1956 1.575 1.260 1.985
3 1957 1.600 1.324 2.119
4 1958 1.803 1.188 2.143
5 1959 1.778 1.079 1.918
6 1960 1.702 1.088 1.852
```

4. Secondary Datasets

1. The second data set is from Kaggle. This ocean acidification data was likely sourced from the Copernicus Marine Service, as the reported values and their table values are very similar.
2. The third dataset is from Wikipedia and covers global surface temperature. This dataset was derived from data presented by the IPCC (Intergovernmental Panel on Climate Change).

5. Data Wrangling

```
library(dplyr)
```

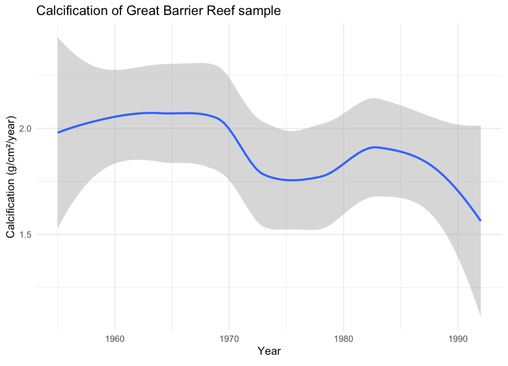
```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(tidyr)
r1 <-read.csv("Desktop/project/ribreef3b.csv")
r2 <-read.csv("Desktop/project/ribreef3a.csv")
r3 <-read.csv("Desktop/project/ribreef.csv")
r4 <-read.csv("Desktop/project/sweetlip2a.csv")
r5 <-read.csv("Desktop/project/sweetlip1b.csv")
r6 <-read.csv("Desktop/project/ribreef3b.csv")
r7 <-read.csv("Desktop/project/kelso2a.csv")
r8 <-read.csv("Desktop/project/kelso2b.csv")
r9 <-read.csv("Desktop/project/sweetlip1a.csv")
r10 <-read.csv("Desktop/project/abrah1h.csv")
coral_files <- rbind(r1,r2,r3,r4,r5,r6,r7,r8,r9,r10)
head(coral_files)
```

```
  Year   Ext   Den  Calc         Name
1 1955 0.889 1.369 1.217 Rib Reef 03B
2 1956 1.397 1.335 1.865 Rib Reef 03B
3 1957 1.829 1.293 2.364 Rib Reef 03B
4 1958 2.032 1.324 2.689 Rib Reef 03B
5 1959 1.626 1.254 2.039 Rib Reef 03B
6 1960 0.660 1.226 0.809 Rib Reef 03B
```

```
library(dplyr)
library(tidyr)
library(ggplot2)
t <- read.csv("Desktop/project/stoneh.csv")
ggplot(t, aes(x = Year, y = Calc)) +
  geom_smooth() +
  labs(title = "Calcification of Great Barrier Reef sample",
       x = "Year", y = "Calcification (g/cm²/year)") +
  theme_minimal()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
acid_data <- read.csv("Desktop/project/oceanacc.csv")
acid_column <- acid_data %>%
  select(Year = year,Ocean_acidification.in_PH.)
coral_and_acid <- coral_files %>%
  left_join(acid_column, by = "Year")
head(coral_and_acid)
```

```
  Year   Ext   Den  Calc         Name Ocean_acidification.in_PH.
1 1955 0.889 1.369 1.217 Rib Reef 03B                        8.15
2 1956 1.397 1.335 1.865 Rib Reef 03B                        8.15
3 1957 1.829 1.293 2.364 Rib Reef 03B                        8.15
4 1958 2.032 1.324 2.689 Rib Reef 03B                        8.15
5 1959 1.626 1.254 2.039 Rib Reef 03B                        8.15
6 1960 0.660 1.226 0.809 Rib Reef 03B                        8.15
```