Insert title of project here

https://github.com/mullja21/Mullens_Chang_Jenkins_872_EDA_Final.git

Irene Chang, Susanna Jenkins, and Jordan Mullens

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Question 1: Was the year-over-year change in public investment in lower-carbon energy RD&D significant during the Obama administration?

2 Dataset Information

The dataset utilized in this analysis was downloaded from International Energy Agency. The IEA's Energy Technology RD&D (Research, Development, & Demonstration) Budgets database provides spending information by energy technology in IEA countries from 1974 to 2022. Data is collected from central or federal government budgets and state-owned companies. Spending categories encompass renewables, nuclear power, fossil fuels, hydrogen, fuel cells, and energy efficiency.

3 Data Wrangling

Downloaded the Energy Technology RD&D (Research, Development, & Demonstration) Budgets. After importing the data to GitHub and Rstudio, the raw dataset was wrangled to suit our analysis. A new data frame was created by extracting the desired years (1985-2015), currency (USD 2021 prices and exchange rates), and economic indications (Total, Fossil Fuels, and Unallocated) for each country. A variable for year-over-year percentage (YoY) difference was created to normalize all comparisons.

```
#create new data frame, change column names
RDD.newcolnames <- RDD.raw %>%
  row_to_names(row_number = 1)

#The column names were imported improperly. The first row contains the correct
#column names, so we used the janitor package since they have a function that
#will replace the column names with the first row.


RDD.countries <-
  filter(RDD.newcolnames,
         Country == "Germany" | Country == "United States",
         Currency == "USD (2021 prices and exchange rates)") %>%
  rename("Economic.Indicators" = "Economic Indicators")


RDD.final <- RDD.countries %>%
  filter(Economic.Indicators == "Total Budget" | Economic.Indicators == "Unallocated " |
  select("Country", "Currency", "Economic.Indicators", "1985", "1986", "1987", "1988", "

RDD.transpose <- data.frame(t(RDD.final))

#In our final data frame, we want each year to have its own row; however,
#in the original dataset, each year was a column. We transposed it.

colnames(RDD.transpose) <- c("Germany.FossilFuels", "Germany.Unallocated", "Germany.Tota

RDD.transpose.rows <- RDD.transpose[!(row.names(RDD.transpose) %in% c("Country", "Curren


RDD.transpose.rows <- RDD.transpose.rows %>%
  mutate_at(1:6, as.numeric)

#creating frame for US Data
RDD.US <- data.frame(Year=c("1985", "1986", "1987", "1988", "1989", "1990", "1991", "199
```

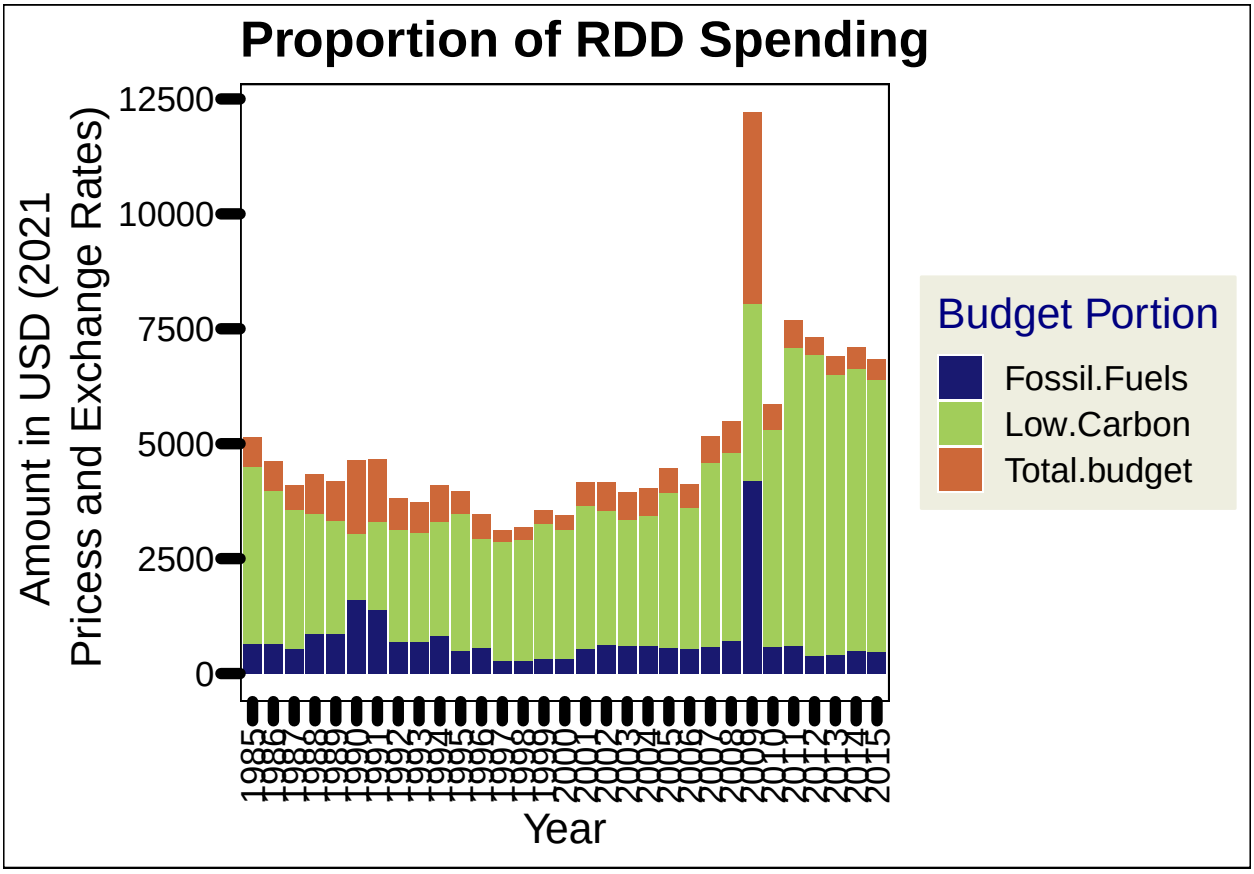
```
                      Fossil.Fuel= RDD.transpose.rows$US.FossilFuels,
                      Unallocated= RDD.transpose.rows$US.Unallocated,
                      Total.Budget= RDD.transpose.rows$US.TotalBudget) %>%
  mutate(Low.Carbon.Energy = c(Total.Budget-Fossil.Fuel-Unallocated)) %>%
  mutate(Percent.Change = ((Low.Carbon.Energy -
                              lag(Low.Carbon.Energy))/lag(Low.Carbon.Energy))*100) %>%
  mutate(Country = "US")


#creating frame for Germany Data
RDD.Germany <- data.frame(Year=c("1985", "1986", "1987", "1988", "1989", "1990", "1991",
                           Fossil.Fuel= RDD.transpose.rows$Germany.FossilFuels,
                           Unallocated= RDD.transpose.rows$Germany.Unallocated,
                           Total.Budget= RDD.transpose.rows$Germany.TotalBudget) %>%
  mutate(Low.Carbon.Energy = Total.Budget-Fossil.Fuel-Unallocated) %>%
  mutate(Percent.Change = ((Low.Carbon.Energy -
                              lag(Low.Carbon.Energy))/lag(Low.Carbon.Energy))*100) %>%
  mutate(Country = "Germany")
```

```
#Adding date column to US data frame and transforming it into date object. Also
#adding a column to denote pre or Obama era for graphing.
RDD.US.Eras <- RDD.US %>%
  mutate(Date = ymd(paste0(RDD.US$Year, "-01-01"))) %>%
  mutate(Era = ifelse(Year <= 2008, "Pre", "Obama"))

#creating one data frame for US and Germany data
US.and.Germany.RDD <- rbind(RDD.Germany, RDD.US)

#Adding Date column to US and Germany dataframe. Formatting so that column is
#transformed to date class

US.and.Germany.RDD <- US.and.Germany.RDD %>%
  mutate(Date = ymd(paste0(US.and.Germany.RDD$Year, "-01-01"))) %>%
  mutate(Era = ifelse(Year <= 2008, "Pre", "Obama"))
```

```
#file for Us and Germany
write.csv(US.and.Germany.RDD, row.names = FALSE, file = "./Processed_Data/US.and.Germany

#File for US alone with eras
write.csv(RDD.US.Eras, row.names = FALSE, file = "./Processed_Data/US.Eras.RDD.csv")
```

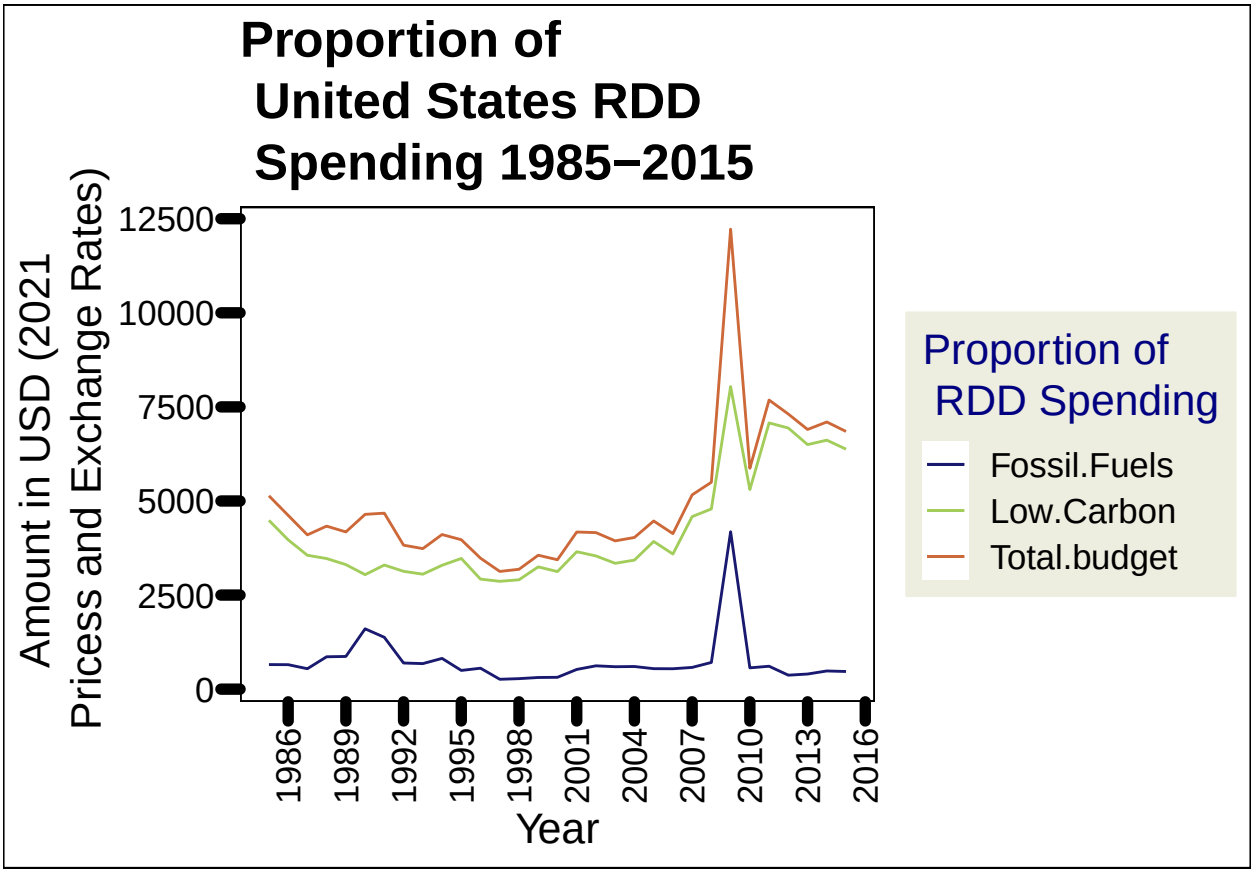
4 Exploratory Analysis

```
# Create the combined bar plot
RDD.US.Eras.Bar.plot <- ggplot(RDD.US.Eras, aes(x = Year)) +
      geom_bar(aes(y = Total.Budget, fill = "Total.budget"), stat= "identity", position
  geom_bar(aes(y = Low.Carbon.Energy, fill = "Low.Carbon"), stat= "identity", position =
  geom_bar(aes(y = Fossil.Fuel, fill = "Fossil.Fuels"), stat= "identity", position = "id
  scale_fill_manual(values = c("Fossil.Fuels" = "midnightblue", "Low.Carbon" = "darkoliv
  labs(fill = "Budget Portion")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  labs(title = "Proportion of RDD Spending",
       y="Amount in USD (2021 \n Pricess and Exchange Rates)",
       x="Year")
print(RDD.US.Eras.Bar.plot)
```




```
#US Line Plot
RDD.US.Eras.Line.plot <- ggplot(RDD.US.Eras, aes(x = Date)) +
  geom_line(aes(y = Total.Budget, color= "Total.budget")) +
  geom_line(aes(y = Fossil.Fuel, color= "Fossil.Fuels")) +
  geom_line(aes(y = Low.Carbon.Energy, color= "Low.Carbon")) +
```

```
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_color_manual(name = "Proportion of \n RDD Spending", values= c(
    "Fossil.Fuels" = "midnightblue",
    "Low.Carbon" = "darkolivegreen3",
    "Total.budget" = "sienna3")) +
  scale_x_date(date_breaks = "3 years", date_labels = "%Y") +
  labs(title = "Proportion of \n United States RDD \n Spending 1985-2015",
       y="Amount in USD (2021 \n Pricess and Exchange Rates)",
       x="Year",
       color= "Legend Title")

print(RDD.US.Eras.Line.plot)
```

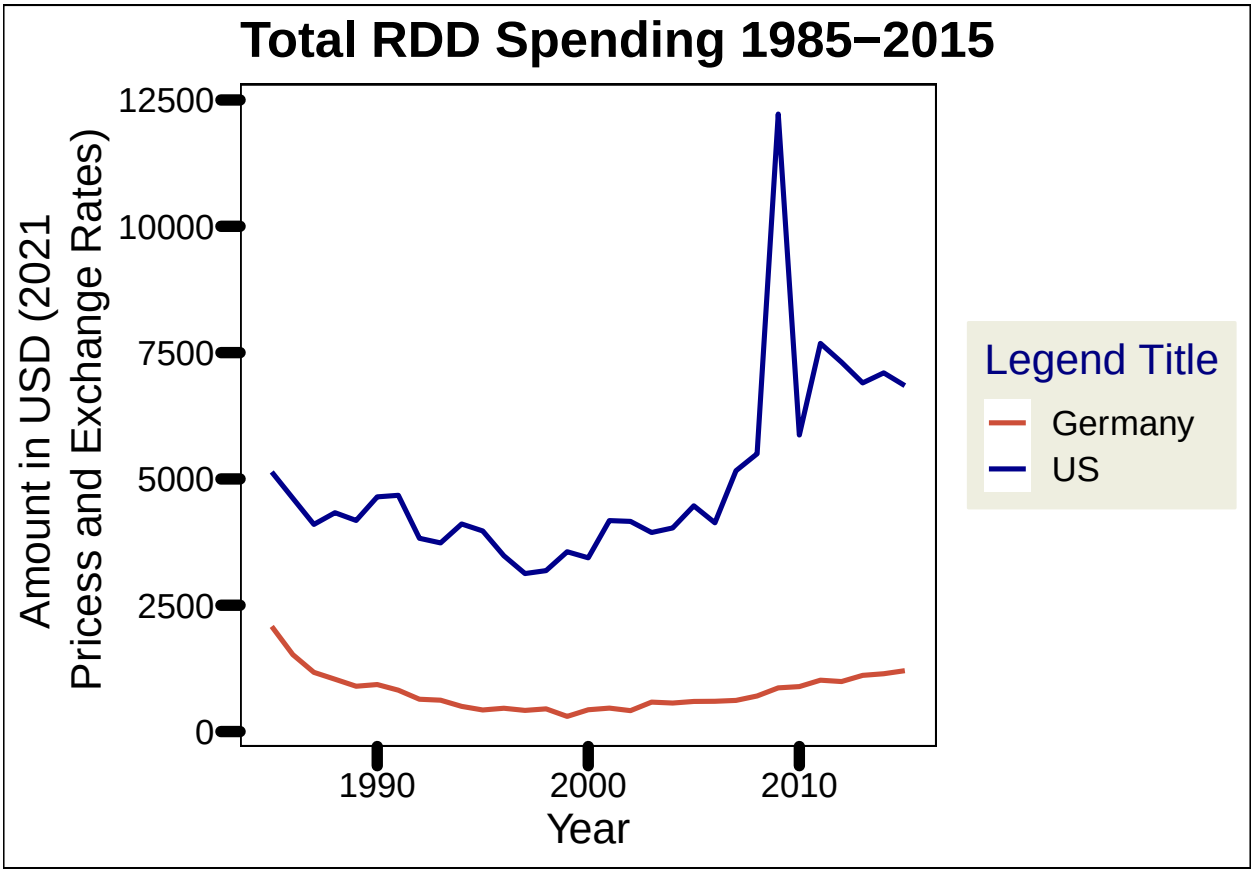



```
# Germ and US Line Plot
US.Germ.total.plot <- ggplot(US.and.Germany.RDD,
                             aes(x = Date,
                                 y = Total.Budget,
                                 color = Country)) +
  geom_line(size= 0.9) +
  scale_color_manual(values = c("tomato3", "darkblue")) +
  labs(title = "Total RDD Spending 1985-2015",
```

```
       y="Amount in USD (2021 \n Pricess and Exchange Rates)",
       x="Year",
       color= "Legend Title")
```

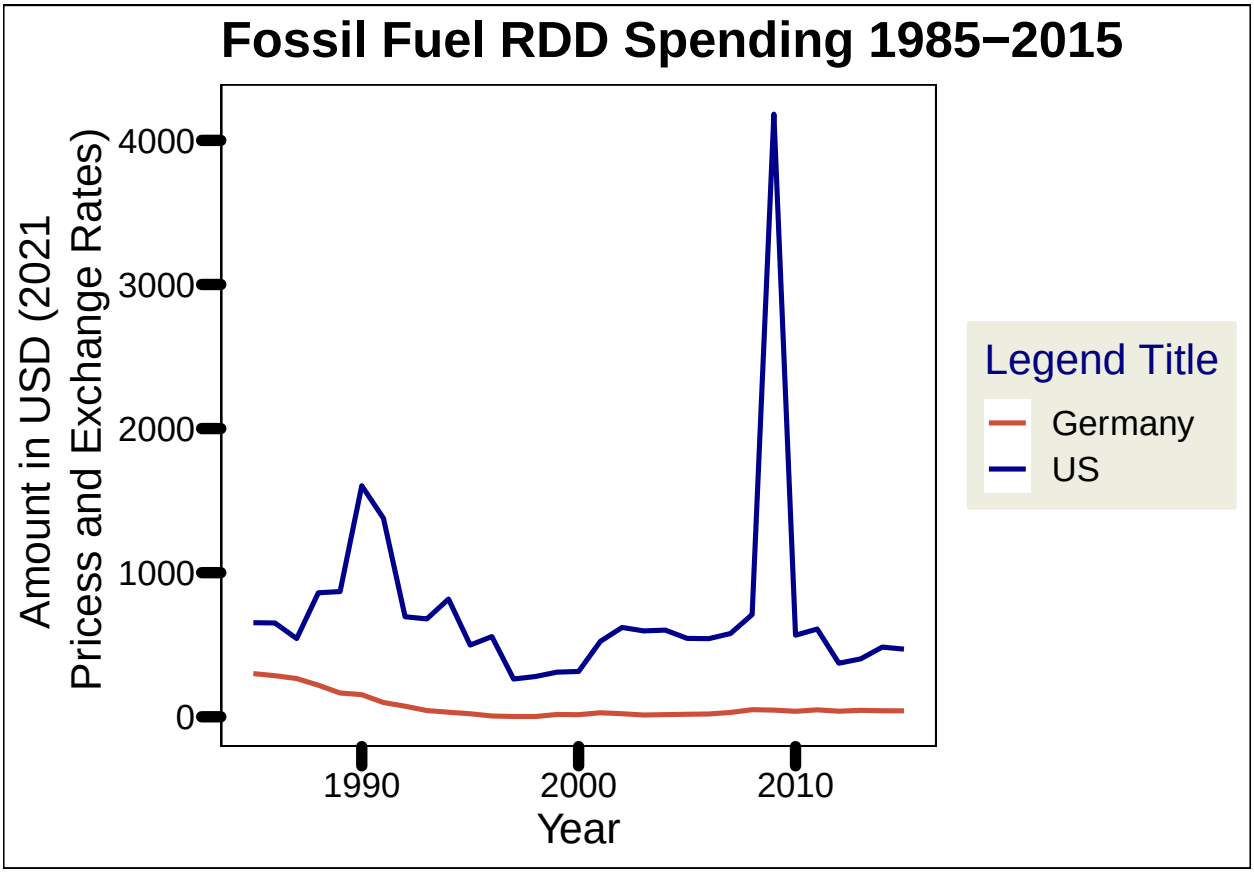
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
```

```
print(US.Germ.total.plot)
```

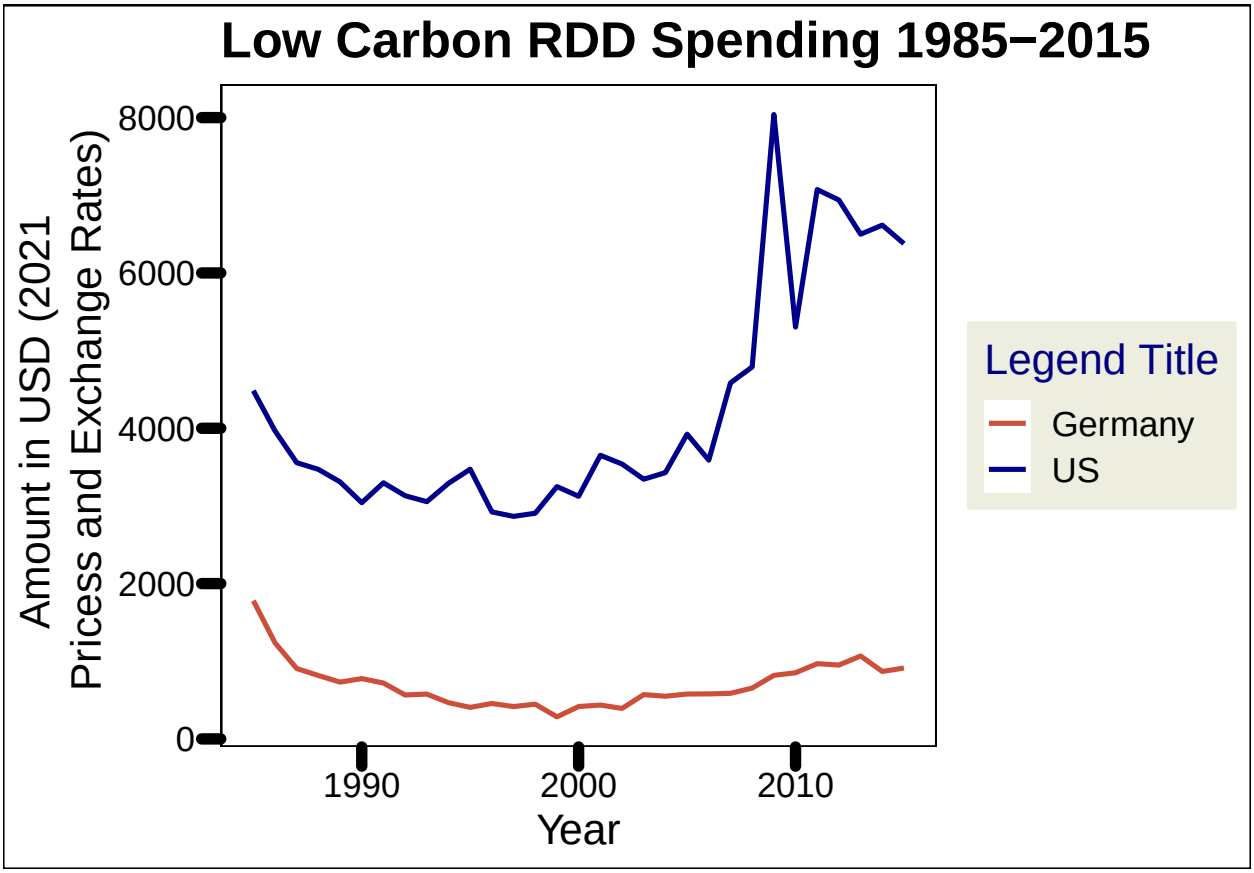



```
#Fossil Fuel Germ and US Line Plot
US.Germ.fossil.plot <- ggplot(US.and.Germany.RDD,
                              aes(x = Date,
                                  y = Fossil.Fuel,
                                  color = Country)) +
  geom_line(size= 0.9) +
  scale_color_manual(values = c("tomato3", "darkblue")) +
  labs(title = "Fossil Fuel RDD Spending 1985-2015",
       y="Amount in USD (2021 \n Pricess and Exchange Rates)",
       x="Year",
       color= "Legend Title")
```

```
print(US.Germ.fossil.plot)
```




```
#Low Carbon Germ and US Line Plot
US.Germ.lowC.plot <- ggplot(US.and.Germany.RDD,
                            aes(x = Date,
                                y = Low.Carbon.Energy,
                                color = Country)) +
  geom_line(size= 0.9) +
  scale_color_manual(values = c("tomato3", "darkblue")) +
  labs(title = "Low Carbon RDD Spending 1985-2015",
       y="Amount in USD (2021 \n Pricess and Exchange Rates)",
       x="Year",
       color= "Legend Title")

print(US.Germ.lowC.plot)
```




```
#Creating a cowplot
library(cowplot)
```

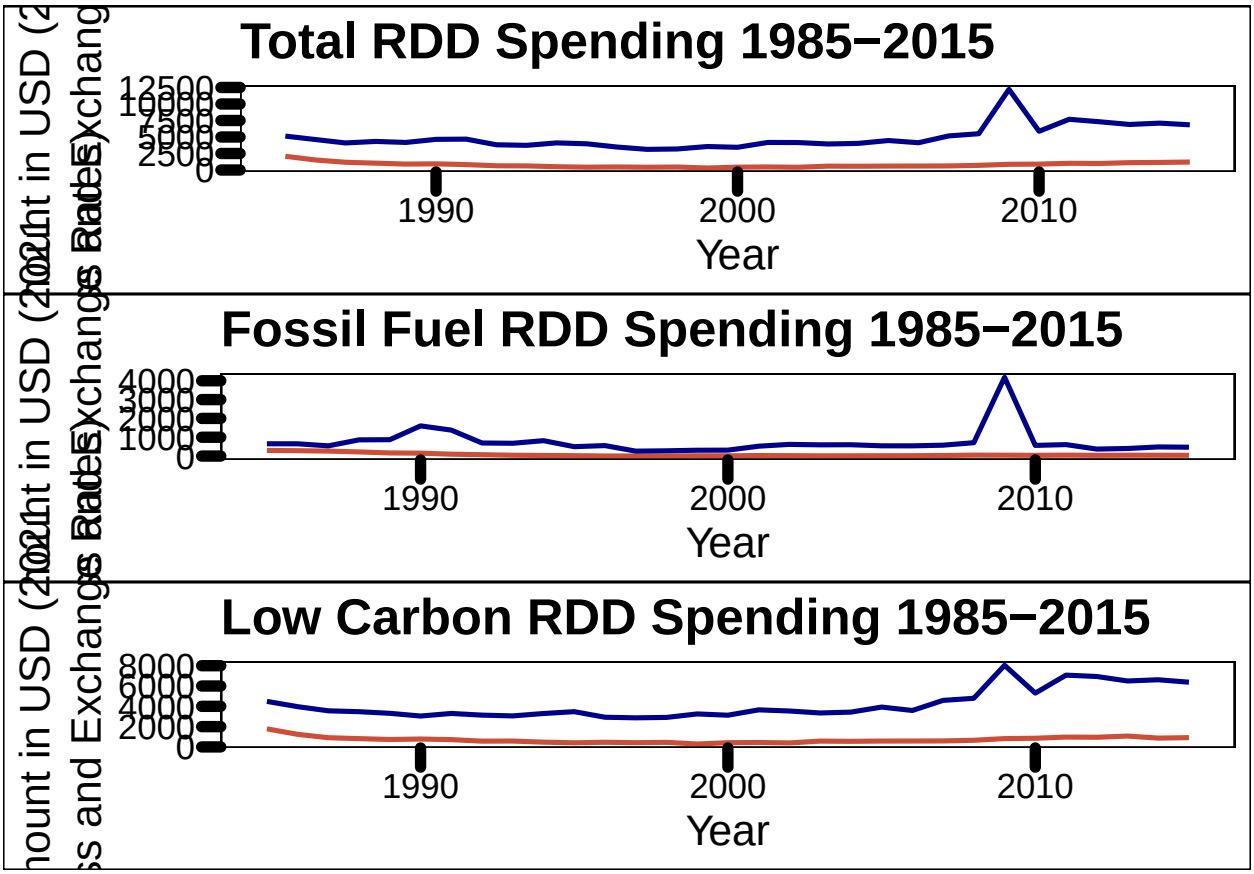
```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggthemes':
##
##     theme_map

## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
legend_cow <- get_legend(
  US.Germ.lowC.plot +
  guides(color = guide_legend(nrow = 1)) +
  theme(legend.position = "right"))
```

```
US.Germ.cowplot <- plot_grid(
  US.Germ.total.plot + theme(legend.position="none"),
  US.Germ.fossil.plot + theme(legend.position="none"),
  US.Germ.lowC.plot + theme(legend.position="none"),
  ncol = 1,
  nrow = 3,
  heights = c(1, 1, 1),
  legend_cow
  )
```

```
## Warning in as_grob.default(plot): Cannot convert object of class numeric into a
## grob.
```

```
print(US.Germ.cowplot)
```



5 Analysis

Our analysis of this dataset focuses on comparing the mean investment YoY difference of the Pre-Obama era and the mean investment YoY difference of the Obama era. However, we need to ensure any changes in investment is unique to the Obama administration and not occurring globally. Therefore, we will also utilize t-tests to compare US investments to Germany's within the same time period.

```
#plotting the pre-Obama data
RDD.US.Eras.plot <- RDD.US.Eras %>%
  ggplot(
         aes(x= Date,
             y= Percent.Change,
             color= Era)) +
  geom_line() +
  geom_point() +
  geom_smooth(data = subset(RDD.US.Eras, Era == "Pre"), method = "lm", se=FALSE, color=
  geom_smooth(data = subset(RDD.US.Eras, Era == "Obama"), method = "lm", se=FALSE, color
  labs(title = "Year-over-Year Percent Change in Public \n Low-Carbon Energy RD&D \n Spe
       y="YoY Percent Change",
       x="Year") +
  scale_color_manual(values = c("darkblue", "darkgreen")) +
  scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
  our_theme +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

print(RDD.US.Eras.plot)
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 1 row containing missing values (`geom_line()`).

## Warning: Removed 1 rows containing missing values (`geom_point()`).
```

```
#plotting the data
US.Germany.RDD.plot <- ggplot(US.and.Germany.RDD,
                              aes(x = Date,
                                  y = Percent.Change,
```

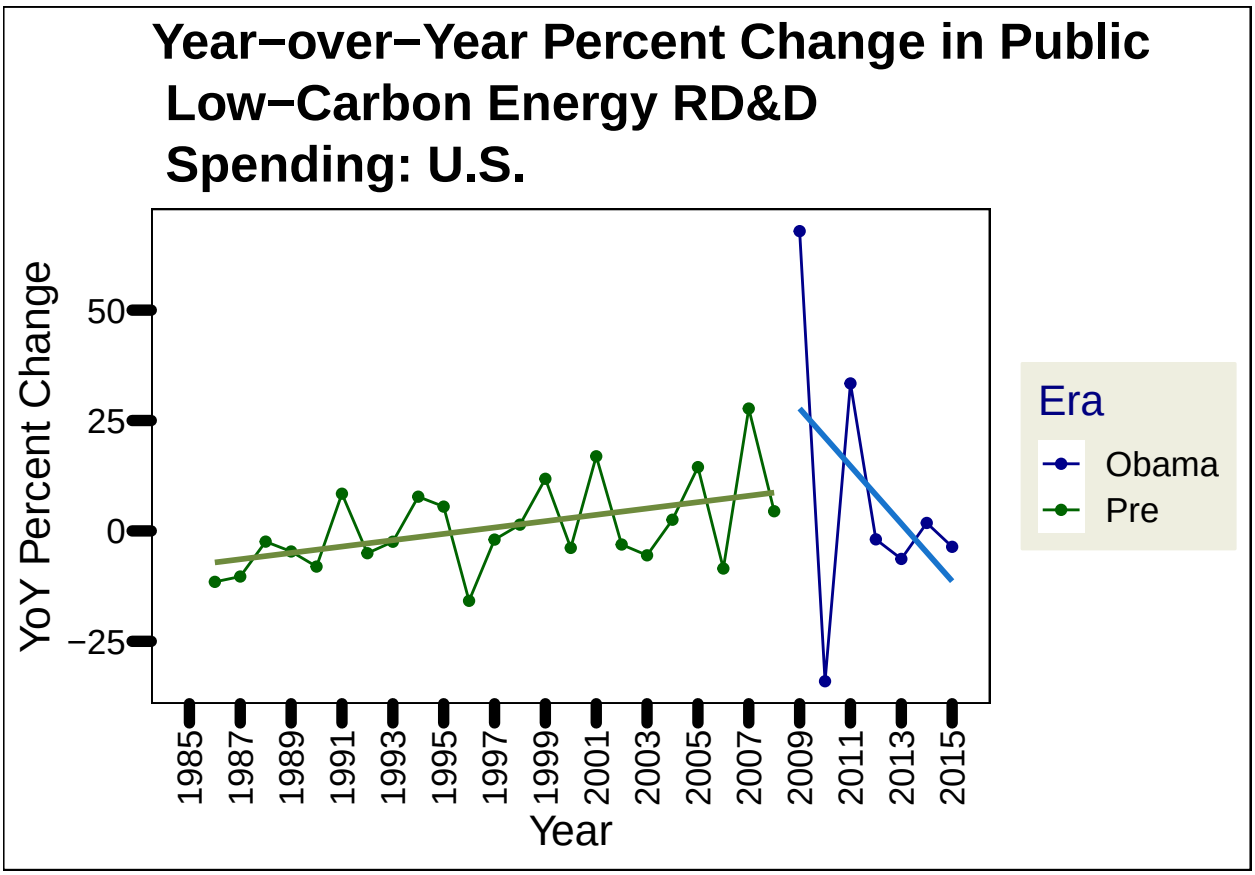


Figure 1: US Year over Year RDD Spending 1985 to 2015

```
                                    color = Country)) +
  geom_line() +
  geom_point() +
  geom_smooth(data = subset(US.and.Germany.RDD, Country == "Germany"), method = "lm", se
  geom_smooth(data = subset(RDD.US.Eras, Country == "US"), method = "lm", se=FALSE, colo
  scale_color_manual(values = c("tomato3", "darkblue")) +
  labs(title = "YoY Change in Public Low-Carbon \n Energy RD&D Spending: \n US vs. Germa
       y="Year-over-Year Percent Change",
       x="Year") +
  scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
  our_theme +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

print(US.Germany.RDD.plot)
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 2 rows containing missing values (`geom_line()`).

## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
# T-test #1) U.S. Public Low-Carbon Energy RD&D Spending: Pre-Obama (1985-2008) vs. du

        #Alternative hypothesis == YoY change in public spending on lower-carbon energ

        #Run two-sample T-test on US RDD "Eras" data to determine if mean YoY change i

US.Pre.v.Post.Obama.ttest <- t.test(RDD.US.Eras$Percent.Change ~
                                      RDD.US.Eras$Era)

US.Pre.v.Post.Obama.ttest
```

```
##
##  Welch Two Sample t-test
##
```

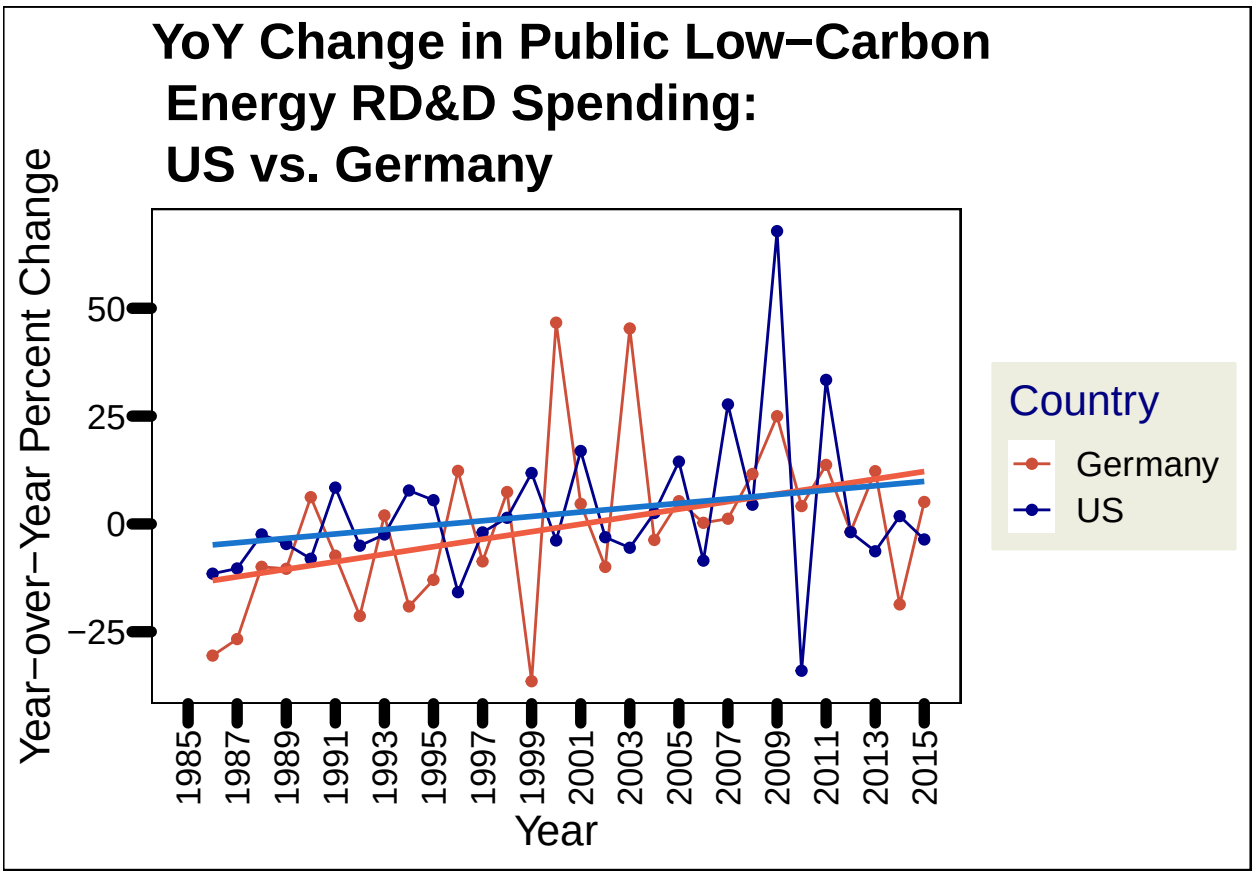


Figure 2: US and Germany Year over Year RDD Spending 1985 to 2015

```
## data:  RDD.US.Eras$Percent.Change by RDD.US.Eras$Era
## t = 0.58747, df = 6.3612, p-value = 0.5771
## alternative hypothesis: true difference in means between group Obama and group Pre is
## 95 percent confidence interval:
##  -23.01113  37.81629
## sample estimates:
## mean in group Obama   mean in group Pre
##           8.1695963           0.7670191
```

```
        #The p-value is 0.5771 which is greater than 0.05 so we must accept our null
```

```
#---------------------------

# T-test #2) Pre-Obama Public Low-Carbon RD&D Spending (1985-2008): U.S. vs Germany

        #Alternative hypothesis == Prior to Obama's inauguration (in the period spann

         ##^ this hypothesis would mean that our null hypothesis is: prior to obama a

Pre.2009.US.Germany.RDD <- filter(US.and.Germany.RDD, Year %in% c(1985:2008))

Pre.2009.US.Germany.ttest <- t.test(Pre.2009.US.Germany.RDD$Percent.Change ~
                                      Pre.2009.US.Germany.RDD$Country)

Pre.2009.US.Germany.ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  Pre.2009.US.Germany.RDD$Percent.Change by Pre.2009.US.Germany.RDD$Country
## t = -0.66411, df = 32.7, p-value = 0.5113
## alternative hypothesis: true difference in means between group Germany and group US i
## 95 percent confidence interval:
##  -12.743123   6.472836
## sample estimates:
## mean in group Germany      mean in group US
##            -2.3681247             0.7670191
```

```
      #The p-value is 0.5113 which is greater than 0.05 so we must accept our null hyp

      #NOTES FROM GOOGLE RE: NULL: If the p-value is below your threshold of significa

#-----------------------------
```

```r
# T-test #3) Obama-era Public Low-carbon RD&D Spending (2009-2015): United States vs.

            ##Alternative Hypothesis == during the Obama presidency (2009-2015), the m

Post.2009.US.Germany.RDD <- filter(US.and.Germany.RDD, Year %in% c(2009:2015))

Post.2009.US.Germany.ttest <- t.test(Post.2009.US.Germany.RDD$Percent.Change ~
                                       Post.2009.US.Germany.RDD$Country)

Post.2009.US.Germany.ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  Post.2009.US.Germany.RDD$Percent.Change by Post.2009.US.Germany.RDD$Country
## t = -0.18605, df = 8.0331, p-value = 0.857
## alternative hypothesis: true difference in means between group Germany and group US i
## 95 percent confidence interval:
##  -33.51499  28.50738
## sample estimates:
## mean in group Germany      mean in group US
##              5.665790              8.169596
```

```r
          #The p-value is 0.857 which is greater than 0.05 so the null hypothesis is a
```

6 Summary and Conclusions

7 References

<add references here if relevant, otherwise delete this section>