Economic Data Analysis Final Project

Indrayan Banerjee

2023-11-25

Exploring Economic Data

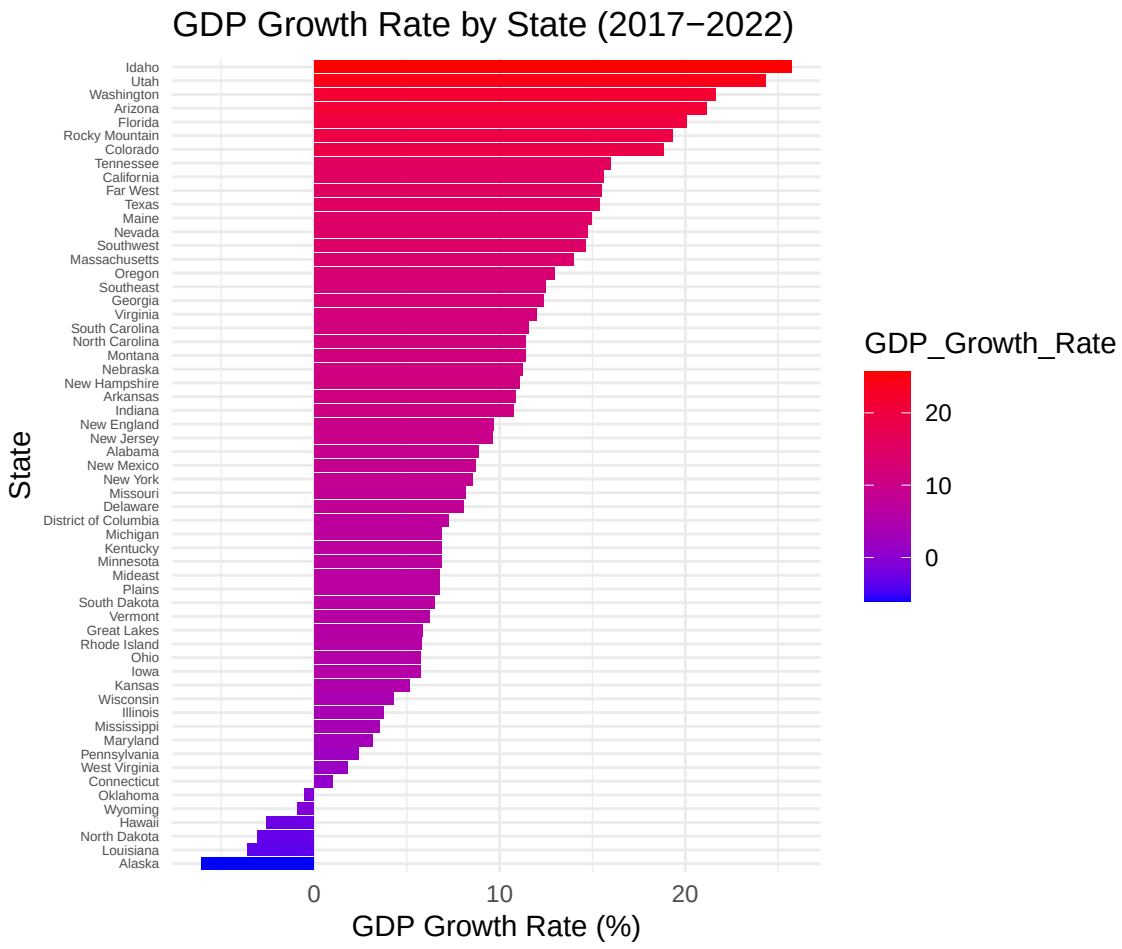
This research will dig deeper into why certain US states and overall regions often grow faster economically than others; according to a recent study by the US Federal Reserve, more and more people move across the country for a new job or more economic Prosperity than ever before. This begs the question: what makes one region or state grow faster than others, and what factors could affect why some states are growing faster by GDP and job growth? Some factors that might come to mind are people moving for lower taxes, better weather, or the cheaper cost of living, but which of these factors directly correlate with job growth or GDP growth of a particular region or state?

To really understand which of these factors impacts how job and GDP growth is happening across the United States, multiple sources must be compiled together after being scraped from their respective websites. To understand the GDP growth rate, the research used data from 2017 to 2022 from the US Federal Reserve on a state level. The state name being used as a key was combined with weather data, average home price, overall tax burden & job growth data for each state. In terms of sources, In addition to the GDP data from the Federal Reserve[^1], tax data from the Tax Foundation[^2], average temperatures[^3], and job growth data were collected from the World Population Review[^4]. Lastly, the median price of a home was collected from Bankrate.com[^5]. All these sources were scraped from their respective websites and combined into one integrated data set. Compiling all these datasets together into an integrated data frame allows this research to have a broad data set to work with, which has multiple factors that can be looked into further.

Data Exploration

Since this is a large data set filled with multiple features it's important to understand the data before we get started. In order to do this I ran a simple data visualization showing GDP growth rate by each individual state just to get a broad picture of what we are studying.

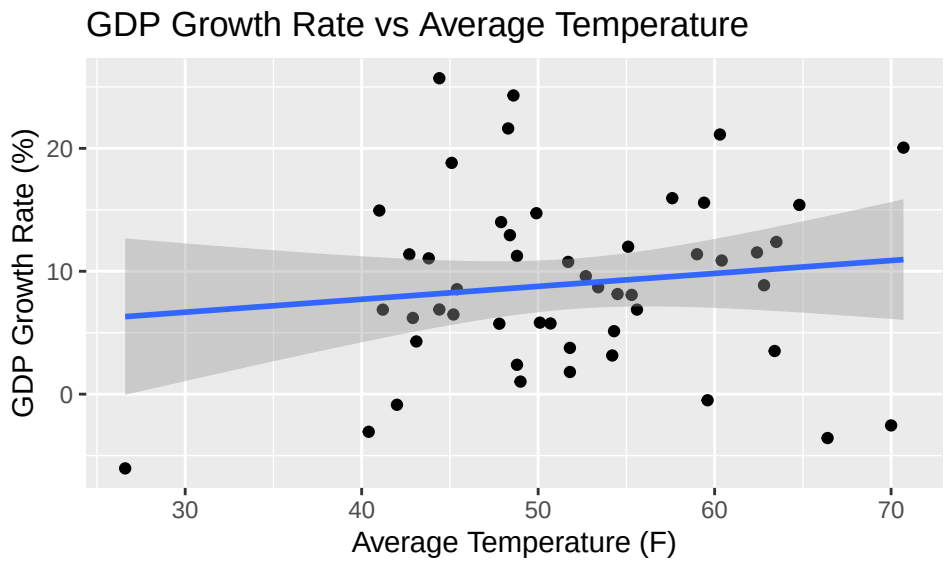
The following figure shows a bar plot showing the GDP growth rate across all 50 states and regional distributions. Broadly, we can see that most of the 50 states grew their GDP during this time; however, we can see five outlier states, Oklahoma, Wyoming, Hawaii, North Dakota, Louisiana, and Alaska, that shrunk their GDP during this period. We can see using the scale that red represents states that were growing the fastest, while blue represents the states that were barely growing or shrinking, such as Alaska.



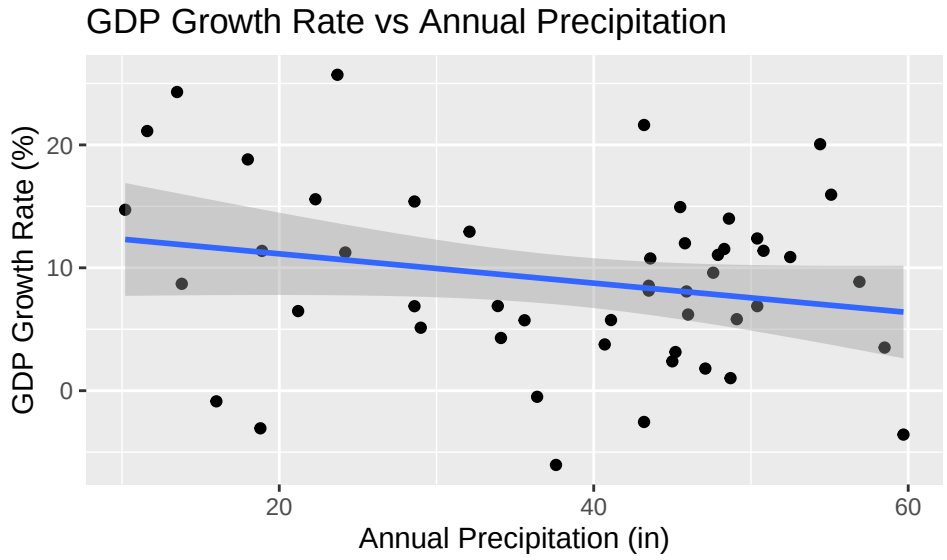
Weather/Climate

To determine what features and factors matter in terms of economics from this dataset, we need to visualize how some factors play out when compared to their correlation with job growth or GDP growth. To understand the weather/climate factor, we can split it into two distinct types: one would measure the correlation between more rainfall and GDP growth, and the other would measure average temperature year-round compared to GDP growth. Both could be factors when people or companies decide to move across the country to a new location; therefore, it is worth studying.

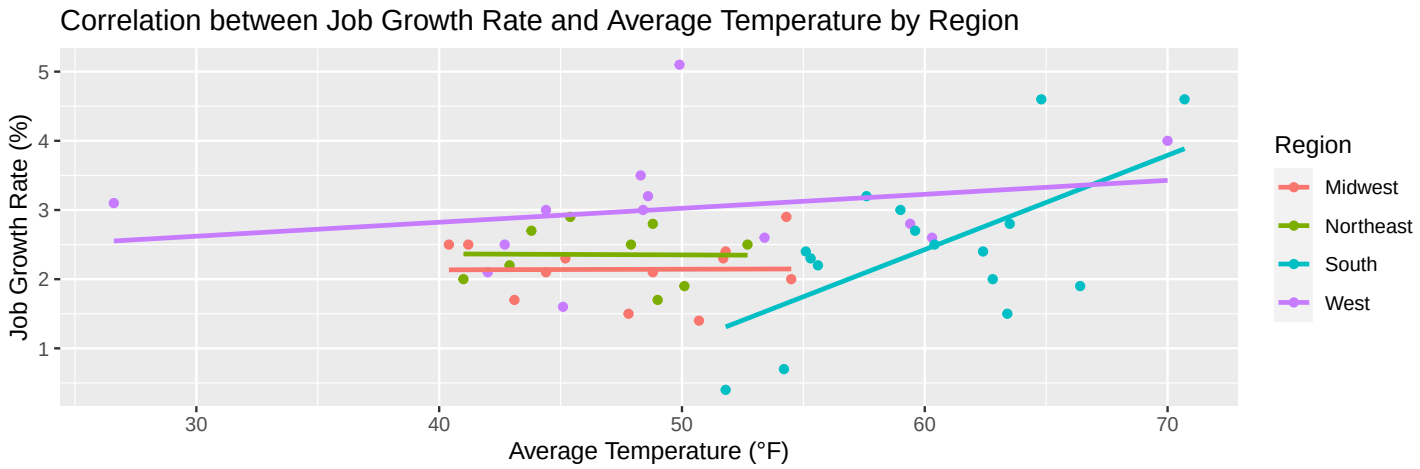
We can visualize this relationship by drawing a scatter plot of states with average temperature values compared to their growth rate. By doing so, we'll understand if the average temperature is a factor that is correlated with GDP growth. 'GDP Growth Rate vs Average Temperature' shows a moderate positive correlation between higher average temperatures and GDP growth. We can infer from this that more and more companies and people are trying to move to warmer or hotter climates rather than colder climates of the country.

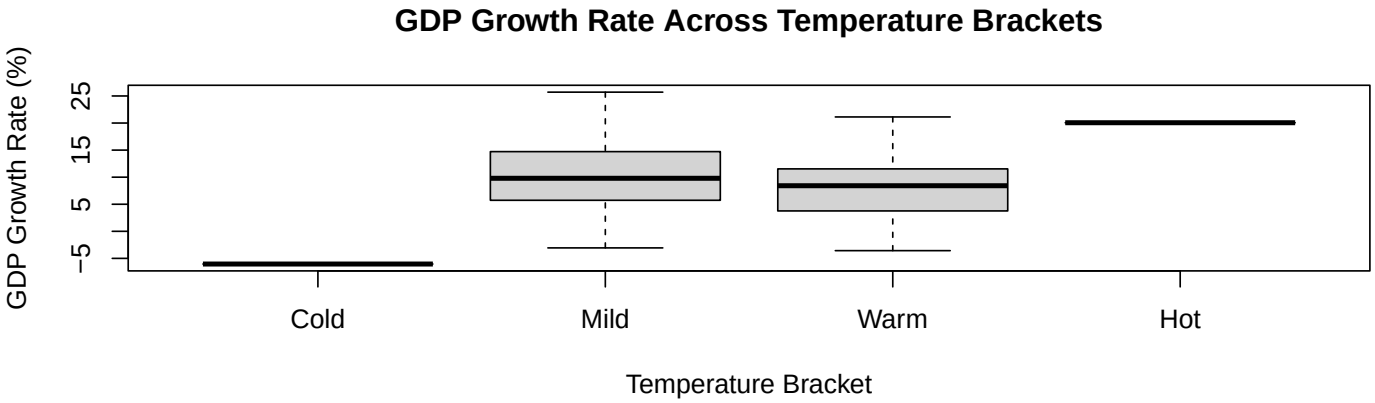


Now, to understand annual precipitation and how it correlates to judging, we can do a similar thing by drawing a scatterplot. We can see if the rainy states are growing faster than non-rainy states or vice versa. In 'GDP Growth Rate vs Annual Precipitation', there's a robust negative correlation between high annual precipitation and high GDP growth. We can infer from this situation that more and more people want to live in drier parts of the country rather than rainy ones, causing more GDP growth in drier areas.



Lastly here, we can see a different type of growth, job growth which for normal people would be the essential thing to look at as that would provide us a picture of where they would thrive economically, AKA where the most jobs are, similar to the other graphs we can see the same trend in which the higher average temperature regions such as the west and the South are Outpacing in job growth as well compared to areas like the Midwest and the Northeast signifying that not only is GDP growing but also jobs.





Here we can also see that hot regions of the country had the highest GDP growth on average, however Mild was close and have an higher upper bound. We can further see from this data that Hot and Mild are growing MUCH faster than Cold regions.

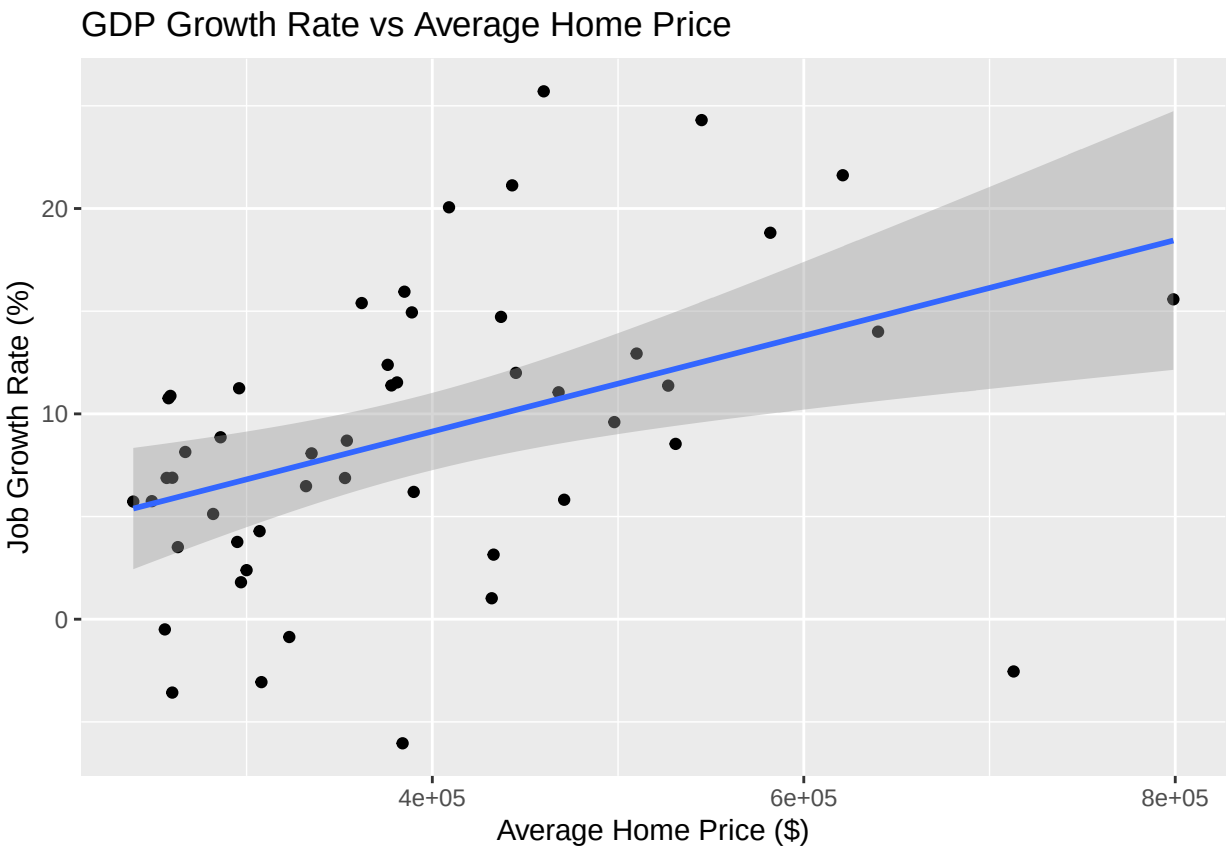
Taxes

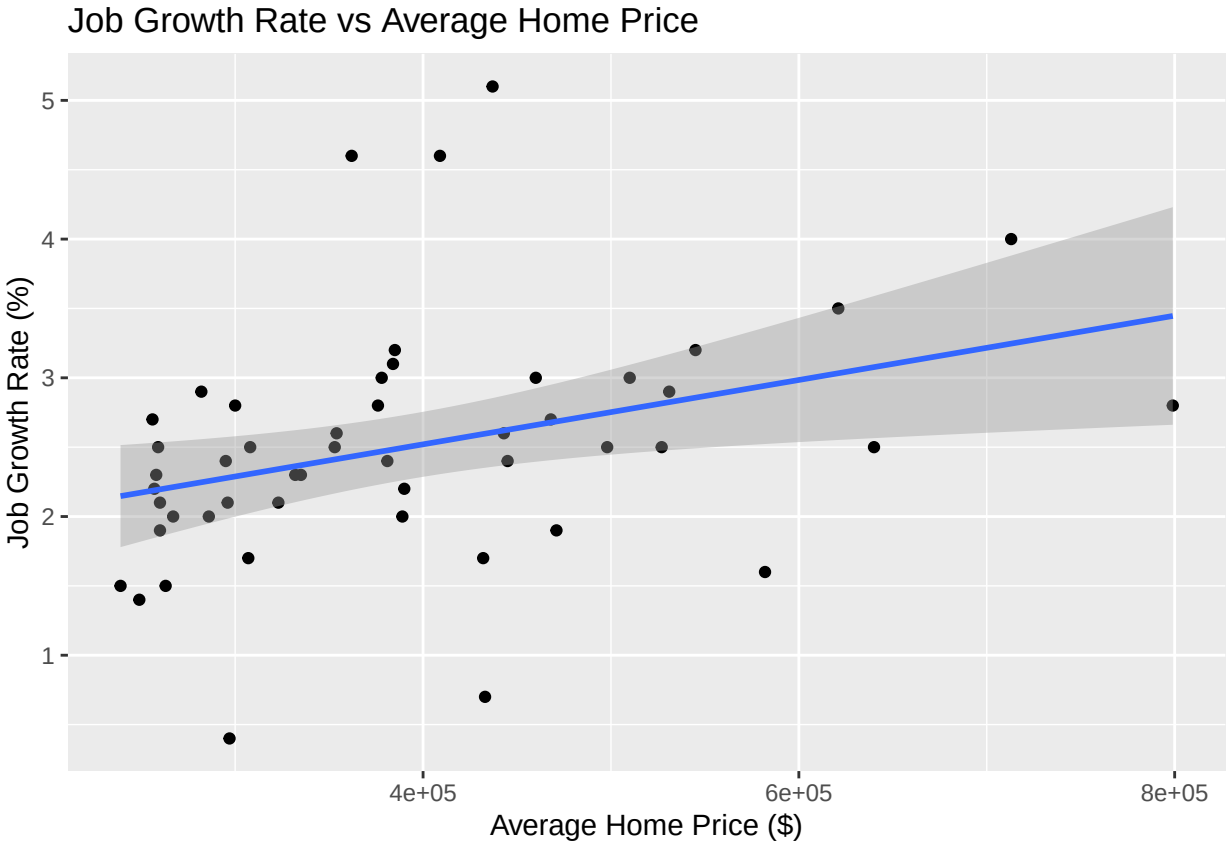
Taxes are also an essential factor that might cause people to move to some areas of the country or certain specific States to determine if the tax and overall tax burden of a particular region or state is having an impact or is in some way correlated with the GDP growth rate we can do so by a similar scatter plot like before however this time not only will we compare states individually but as regions itself. This is because certain parts of the country often have similar identities and with that comes similar taxes, for example, the Northeast is known for its high taxes however the Midwest and the South is known for a lower amount of taxes by recently dividing up the country we can determine if this type of situation is having a regional impact on the regional economy.

In this chart we can clearly see a trend in which states that have overall tax burdens between 8 to 12% have much higher GDP growth rates than ones above 12% or below 8%. the signifies that not only is there too much tax but also signifies that states that might have lower taxes might not be keeping up with things like infrastructure which might not cause people or companies to want to move there therefore. Though these are statistical inferences some of the inferences that can be made from this data can be seen in the real world for example, States like Alaska which are considered to be very tax-friendly and often have lower tax burdens grow much slower than States like Washington which have higher tax burdens however have been investing in the infrastructure and have been bringing in companies to their state and therefore having a higher amount of GDP growth rate.

Home Price

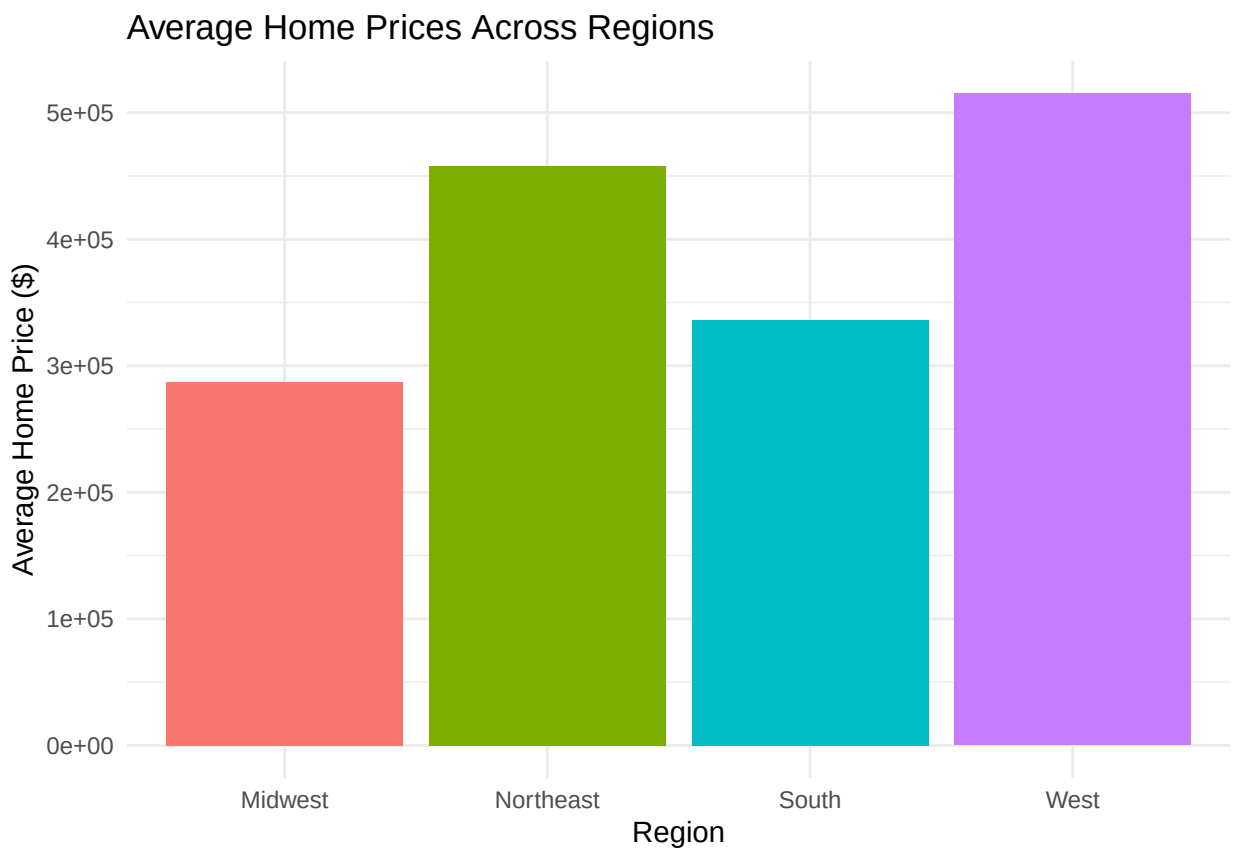
Out of the many things that people and companies consider when moving to a new region or part of the country is the cost of living, and a majority of that comes from housing or the cost of Office Space; for ordinary people, they will be looking at things like the average home price to determine if it is economically significant enough to move to a new part of the country. Many hypothesize that lower average home prices would mean that more people and companies want to move into those regions and, therefore, would significantly increase their GDP growth rates however, to visualize this and see if this trend actually exists, we can create a scatter plot to determine if we see a positive or negative correlation





By looking back to the first chart we looked at in this research, we saw that the West and Far West regions of the country actually had the highest GDP growth rates, and now, since we see in this chart that higher average home prices are positively correlated with not only higher job growth but as well as higher GDP growth we can hypothesize by combining our knowledge

from both charts that the western region of the country is probably the most expensive place to live in as it has the highest amount of GDP growth rate.



As we can see from this chart, we're correct. The western part of the country is the most expensive part of the country to live in. Since we remember, it is also the part of the country that's growing the fastest. By using data analytics and data visualization, we were able to predict correctly something that would have been hard to understand without this type of data, showing how vital this type of research could be in economics.

Region	Average Home Price Per Region
West	$515,231.0000
Northeast	$457,667.0000
South	$336,375.0000
Midwest	$287,167.0000

In the table above we can see that the difference in average home price across the Northeast and West is still a solid $57,564. Showing that the western US is by far has the most expensive real estate market in the country.

Findings and Conclusions

After considering factors like taxes, average temperatures, average precipitation, average home cost, and job growth, we can determine that the most critical factors that were correlated with positive job growth and GDP growth were higher temperatures, lower amounts of precipitation, and a medium amount of tax burden. by understanding this type of economic data, further research can be done to predict how states can improve their chances of growing their GDP faster and bringing more companies and jobs with them.

Code Appendix

```r
#Libraries
library(dplyr)
library(rvest)
library(ggplot2)

#Load GDP Data from St. Louis Fed
gdp_data <- read.csv("gdpdata.csv") %>%
  mutate(across(starts_with("X"), ~ as.numeric(gsub(",", "", .)))) %>%
  mutate(GDP_Growth_Rate = ((X2022 - X2017) / X2017) * 100) %>%
  filter(GeoName != "United States") %>%
  select(State = GeoName, GDP_Growth_Rate)

# Scrape Tax Rate data
tax_url <- "https://taxfoundation.org/data/all/state/tax-burden-by-state-2022/"
tax_data <- read_html(tax_url) %>%
  html_table(fill = TRUE) %>%
  .[[1]] %>%
  rename(State = 1, Tax_Rate = 2) %>%
  mutate(State = as.character(State), Tax_Rate = as.numeric(gsub("%", "", Tax_Rate)))

# Scrape Weather data
weather_url <- "https://worldpopulationreview.com/state-rankings/average-temperatures-by-state"
weather_data <- read_html(weather_url) %>%
  html_table(fill = TRUE) %>%
  .[[1]] %>%
  rename(State = 1, Avg_Temperature = 2) %>%
  mutate(State = as.character(State), Avg_Temperature = as.numeric(gsub("[^0-9\\.]", "", Avg_Temperature)))

# Scrape Employment data
employment_url <- "https://worldpopulationreview.com/state-rankings/job-growth-by-state"
employment_data <- read_html(employment_url) %>%
  html_table(fill = TRUE) %>%
  .[[1]] %>%
  rename(State = 1, Job_Growth_Rate = 4) %>%
  mutate(State = as.character(State), Job_Growth_Rate = as.numeric(gsub("[^0-9\\-\\.]", "", Job_Growth_Rate)))



# Scrape Home Price data
home_price_url <- "https://www.bankrate.com/real-estate/median-home-price/#median-price-by-state"
home_price_data <- read_html(home_price_url) %>%
  html_table(fill = TRUE) %>%
  .[[1]] %>%
  rename(State = 1, Avg_Home_Price = 2) %>%
  mutate(State = as.character(State), Avg_Home_Price = as.numeric(gsub("[^0-9]", "", Avg_Home_Price)))



#Adding in the tax, weather and job data
combined_data <- gdp_data %>%
  left_join(tax_data, by = "State") %>%
  left_join(weather_data, by = "State") %>%
  left_join(employment_data, by = "State") %>%
  filter(!is.na(Tax_Rate) & !is.na(Rank) & !is.na(Avg_Temperature))

#Adding in the home price data
combined_data <- combined_data %>%
  left_join(home_price_data, by = "State")
```

```r
#Defining US regions by state
map_region_division <- function(state) {
  northeast_states <- c("Connecticut", "Maine", "Massachusetts", "New Hampshire", "Rhode Island", "Vermont", "
  midwest_states <- c("Illinois", "Indiana", "Michigan", "Ohio", "Wisconsin", "Iowa", "Kansas", "Minnesota", "
  south_states <- c("Delaware", "Florida", "Georgia", "Maryland", "North Carolina", "South Carolina", "Virgini
  west_states <- c("Arizona", "Colorado", "Idaho", "Montana", "Nevada", "New Mexico", "Utah", "Wyoming", "Alas

  if (state %in% northeast_states) {
    return("Northeast")
  } else if (state %in% midwest_states) {
    return("Midwest")
  } else if (state %in% south_states) {
    return("South")
  } else if (state %in% west_states) {
    return("West")
  } else {
    return("Unknown")
  }
}

#Get combined data with the new regions created
combined_data$Region <- sapply(combined_data$State, map_region_division)

#Visual on GDP Growth Rate by State (2017-2022)
ggplot(gdp_data, aes(x = reorder(State, GDP_Growth_Rate), y = GDP_Growth_Rate, fill = GDP_Growth_Rate)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "GDP Growth Rate by State (2017-2022)",
       x = "State",
       y = "GDP Growth Rate (%)") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 5 ,hjust = 1)) +
  scale_fill_gradient(low = "blue", high = "red")

#Visual on GDP Growth Rate vs Average Temperature
ggplot(combined_data, aes(x = Avg_Temperature, y = GDP_Growth_Rate)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(title = "GDP Growth Rate vs Average Temperature", x = "Average Temperature (F)", y = "GDP Growth Rate (

#Visual on GDP Growth Rate vs Annual Precipitation
ggplot(combined_data, aes(x = `Avg. Annual Precipitation (in)`, y = GDP_Growth_Rate)) +

  geom_point() +

  geom_smooth(method = lm) +

  labs(title = "GDP Growth Rate vs Annual Precipitation", x = "Annual Precipitation (in)", y = "GDP Growth Rat

#Visual on Correlation between Job Growth Rate and Average Temperature by Region
ggplot(data = combined_data, aes(x = Avg_Temperature, y = Job_Growth_Rate, color = Region)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(x = "Average Temperature (°F)", y = "Job Growth Rate (%)") +
  ggtitle("Correlation between Job Growth Rate and Average Temperature by Region")

#Creating divisions widthin the Avg Temperatures
combined_data$Temp_Bracket <- cut(combined_data$Avg_Temperature, breaks = c(-Inf, 32, 50, 70, Inf), labels = c
boxplot(GDP_Growth_Rate ~ Temp_Bracket, data = combined_data, main = "GDP Growth Rate Across Temperature Brack
```

```r
#Visual on GDP Growth Rate vs Tax Rate by Region
ggplot(data = combined_data, aes(x = Tax_Rate, y = GDP_Growth_Rate)) +
  geom_point(aes(color = Region)) +
  geom_smooth(method = lm) +
  labs(title = "GDP Growth Rate vs Tax Rate by Region", x = "Overall Tax Rate (%)", y = "GDP Growth Rate (%)")

#Visual on Job Growth Rate vs Avg Home Price
ggplot(combined_data, aes(x = Avg_Home_Price, y = GDP_Growth_Rate)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(title = "GDP Growth Rate vs Average Home Price", x = "Average Home Price ($)", y = "Job Growth Rate (%)


#Visual on Job Growth Rate vs Average Home Price
ggplot(combined_data, aes(x = Avg_Home_Price, y = Job_Growth_Rate)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(title = "Job Growth Rate vs Average Home Price", x = "Average Home Price ($)", y = "Job Growth Rate (%)

#Getting regional Data
regional_data <- combined_data %>%
  group_by(Region) %>%
  summarise(Avg_GDP_Growth = mean(GDP_Growth_Rate, na.rm = TRUE),
            Avg_Tax_Rate = mean(Tax_Rate, na.rm = TRUE),
            Avg_Job_Growth = mean(Job_Growth_Rate, na.rm = TRUE),
            Avg_Home_Price = mean(Avg_Home_Price, na.rm = TRUE))


#Getting Avg Home Price by Region
avg_home_price_by_region <- combined_data %>%
  group_by(Region) %>%
  summarise(Avg_Home_Price = mean(Avg_Home_Price, na.rm = TRUE)) %>%
  arrange(desc(Avg_Home_Price))

#Visual about Average Home Prices Across Regions
ggplot(avg_home_price_by_region, aes(x = Region, y = Avg_Home_Price, fill = Region)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Home Prices Across Regions", x = "Region", y = "Average Home Price ($)") +
  theme_minimal() +
  theme(legend.position = "none")
```

References

[^1]: St. Louis Federal Reserve. (2023). GDP Growth Rate By State.

[^2]: Tax Foundation. (2023). Overall Tax Burden By State.

[^3]: World Population Review. (2023). Average Temperatures by State.

[^4]: World Population Review. (2023). Year-Over-Year Job Growth Rate By State.

[^5]: Bankrate.com. (2023). Average Home Price By State.