

**Project Report**  
**Josh Anderson, Christine Clutter, Ryan Lang, Connor Stevenson**  
**US Unemployment and Netflix Subscriptions**

## **1. Introduction**

There are many factors that contribute to the unemployment rate increasing or decreasing. For example, the housing market recession and the COVID-19 pandemic caused major spikes in the unemployment rate. However, we were curious what kind of effect the unemployment rate could have on something else. When a person is unemployed, one would assume that a person would spend some of their new free time watching more television. Due to this, streaming services such as Netflix would have much to gain from increased unemployment

In this project, we utilized data on the unemployment rate since 2001 from the US Bureau of Labor Statistics and web-scraped Netflix data since 2011 from [businessofapps.com](https://businessofapps.com) to discover certain changes in the unemployment rate and Netflix subscriptions over time. We will also determine what kind of effect the unemployment rate has on Netflix subscriptions.

## **2. Data**

This project uses two primary sources of data: BusinessofApps data on Netflix subscriptions from the past decade, and US Bureau of Labor Statistics data on the United States unemployment rate since 2001.

### *2.1 Netflix Subscriptions*

To identify and collect data on Netflix Subscriptions, we consulted the Business of Apps article on Netflix revenue and usage statistics. This page contains a section titled “Netflix users (paying subscribers) by quarter” that has data on the total amount of Netflix subscribers since 2011 and lists the data on a quarterly basis.

From the “Netflix users by quarter” data table, we were able to scrape and collect the Quarter and Netflix users in millions. This process produced a data frame with 38 quarters that contained the total amount of Netflix subscribers. The scraping code and all the graphs in the analysis are included in the R script “**Final Wrangling Project.R**”

Another table that was webscraped, but not included in the final merge was Netflix users by region 2011-2018. This data table was used to make other statistical analysis related to the final merged data.

**Netflix site:** <https://www.businessofapps.com/data/netflix-statistics/>

### *2.2 US unemployment*

The US Bureau of Labor Statistics dataset on US unemployment is a collection of the US unemployment rate since April 2001.

The US Bureau of Labor Statistics divides the unemployment data into 8 different categories. The 8 categories are total unemployment rate, men 20 years and over, women 20 years and over, 16 to 19 years old, white, black or African-American, Asian, and Hispanic or Latino.

In order to merge this dataset with the Netflix subscribers dataset, we changed the month column name to year, and took data from every three months so it was quarterly. For example, a cell would read “Q3 2011”, which matches the format of the Netflix subscribers Year column.

Unemployment site: <https://data.bls.gov/pdq/SurveyOutputServlet>

### 2.3 Merged Data Table Features

This table describes the contents of the merged unemployment and Netflix data. All data feature names, data type and description of the data features are below.

**Table 1 Data dictionary**

Column	Type	Description
<b>Year</b>	<b>numeric</b>	<b>Given year</b>
<b>Quarter</b>	<b>text</b>	<b>Quarter of the year</b>
<b>Subscribers in Millions</b>	<b>numeric</b>	<b>Number of total Netflix subscribers at a given time</b>
<b>Total</b>	<b>numeric</b>	<b>Total unemployment percentage of the United States</b>
<b>Men 20 years and over</b>	<b>numeric</b>	<b>Unemployment percentage of men over the age of 20</b>
<b>Women 20 years and over</b>	<b>numeric</b>	<b>Unemployment percentage of women over the age of 20</b>
<b>16 to 19 years old</b>	<b>numeric</b>	<b>Unemployment percentage of people between 16 and 19 years old</b>
<b>White</b>	<b>numeric</b>	<b>Unemployment percentage of the white population in America</b>
<b>Black or African American</b>	<b>numeric</b>	<b>Unemployment percentage of the black or African-American population in America</b>
<b>Asian</b>	<b>numeric</b>	<b>Unemployment percentage of the Asian population in America</b>
<b>Hispanic or Latino</b>	<b>numeric</b>	<b>Unemployment percentage of the hispanic or latino population in America</b>

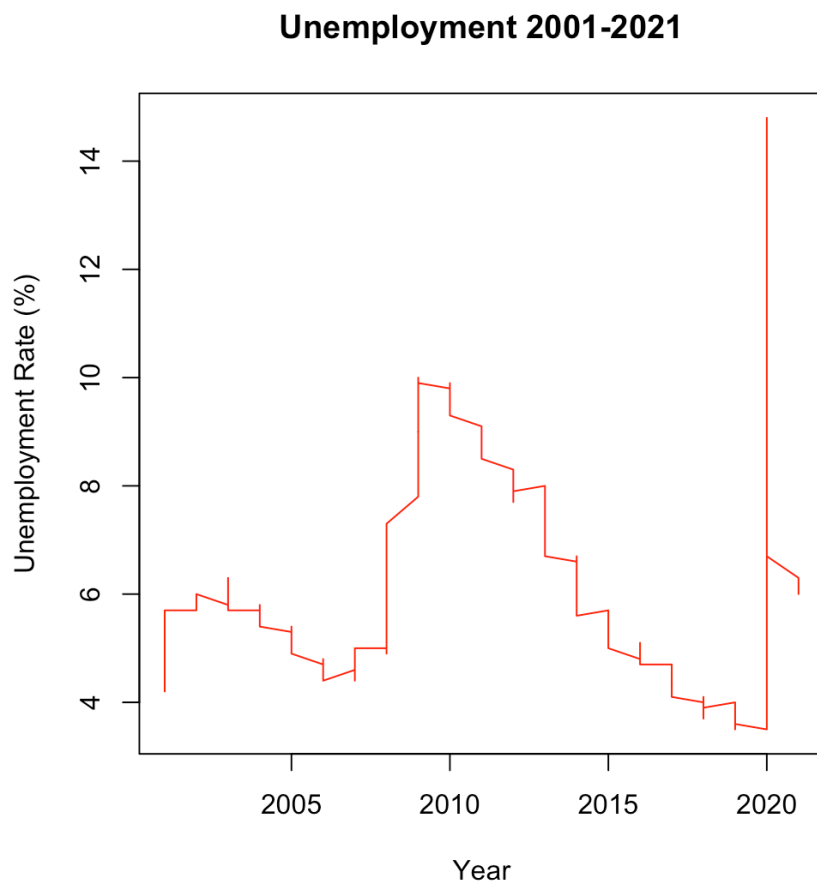
### 3. Analysis

The goal of this project is to examine how the US unemployment rate has changed

through the years, also how the unemployment rate has affected the total number of Netflix subscriptions.

### 3.1 US Unemployment Rate from 2001-2021

How has the US unemployment rate changed over the course of the 21st century? We created a line chart using the plot() function. Using the Year column of data as the x-axis and the Unemployment Rate column data as the y-axis, we can visualize the transformation of the unemployment rate from 2001-2021.



The line chart above indicates that the United States civilian unemployment rate has gone through numerous ups and downs over the past two decades. Between 2001 and 2007, the economy was riding off the high of the dot com boom. Due to the successful economy, the unemployment rate decreased until 2007. The first major change in unemployment began at the end of 2007 with the housing market crash. From then until 2009, the unemployment rate rapidly increased.

After the recession, the unemployment rate gradually decreased until it reached its lowest point in almost 50 years. Finally, the major spike in the unemployment rate can be attributed to the COVID-19 pandemic. Many states went into shutdown causing many small businesses to close, resulting in many US civilians jobless.

### Yearly Unemployment Summary Table

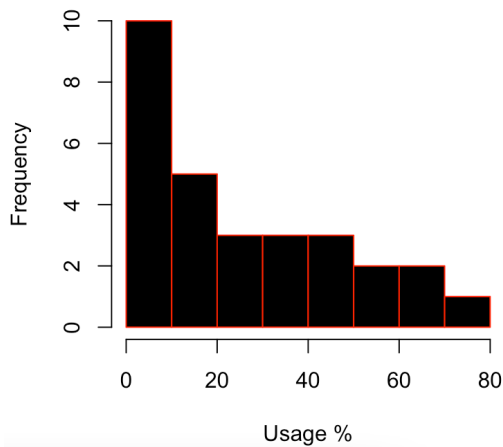
Year	Min	Max	Mean
2001	4.2	5.7	4.741667
2002	5.7	6.0	5.783333
2003	5.7	6.3	5.991667
2004	5.4	5.8	5.541667
2005	4.9	5.4	5.083333
2006	4.4	4.8	4.608333
2007	4.4	5.0	4.616667
2008	4.9	7.3	5.800000
2009	7.8	10.0	9.283333
2010	9.3	9.9	9.608333
2011	8.5	9.1	8.933333
2012	7.7	8.3	8.075000
2013	6.7	8.0	7.358333
2014	5.6	6.7	6.158333
2015	5.0	5.7	5.275000
2016	4.7	5.1	4.875000
2017	4.1	4.7	4.350000
2018	3.7	4.1	3.891667
2019	3.5	4.0	3.683333
2020	3.5	14.8	8.108333
2021	6.0	6.3	6.150000

To summarize United States unemployment fluctuations, this chart shows the min, max and mean of each year from 2001-2021. As shown, the highest recorded unemployment rate in this 20 year span was during 2020 (14.8%), while the lowest was in 2019 and 2020 (3.5%). The year with the highest unemployment rate was 2009 (9.28%).

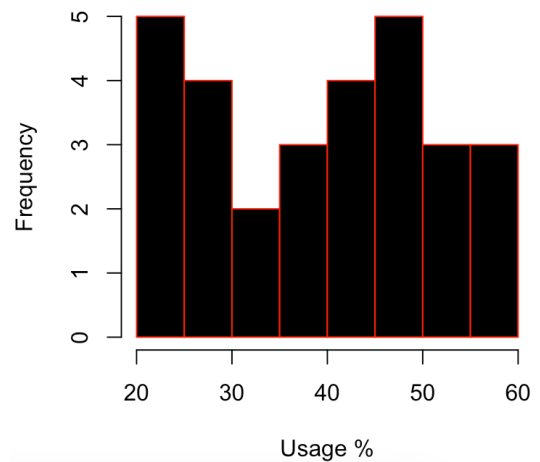
### 3.2 Domestic vs. International Netflix Usage

How does the United States Average Netflix usage compare to the International Average from 2011-2018? Two histograms were created using the hist() function. The first displays the average Netflix usage percentage for international citizens and the second displays the average usage percentage of Netflix for American citizens.

**International Netflix Usage 2011-2018**



**U.S. Netflix Usage 2011-2018**

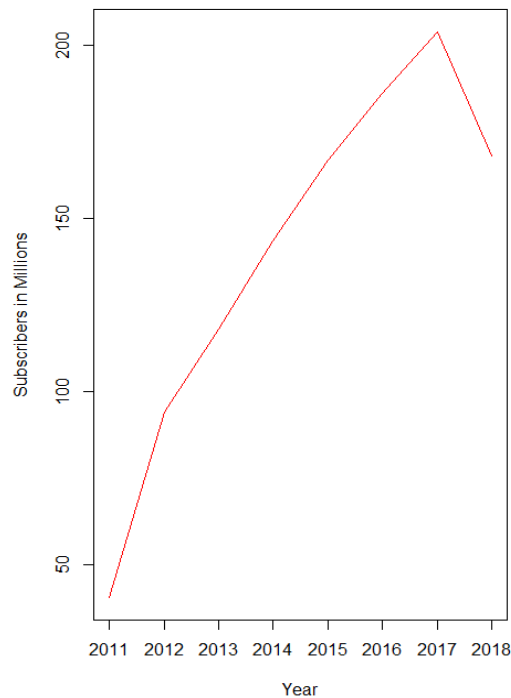


The first histogram shows the reluctance of international users to subscribe to Netflix. From 2011-2015, their usage percentage was below 20%. Then from 2015-2016 you see a gradual climb until you get to 2017. From 2017-2018, there is a large spike in usage with the percentage going from 45-73%.

The second histogram shows how the United States hopped on the Netflix bandwagon earlier, but tapered off as the years passed. From 2011-2013, usage in the U.S. was between 20-30%. From 2013-2018, there was a steady climb from 30-57%. Netflix has always been popular in the United States, but with the addition of other popular streaming services such as Hulu, Hbo Go, and Amazon Video subscribers have started to take their money elsewhere. This is part of why the growth of Netflix has tapered off in the United States.

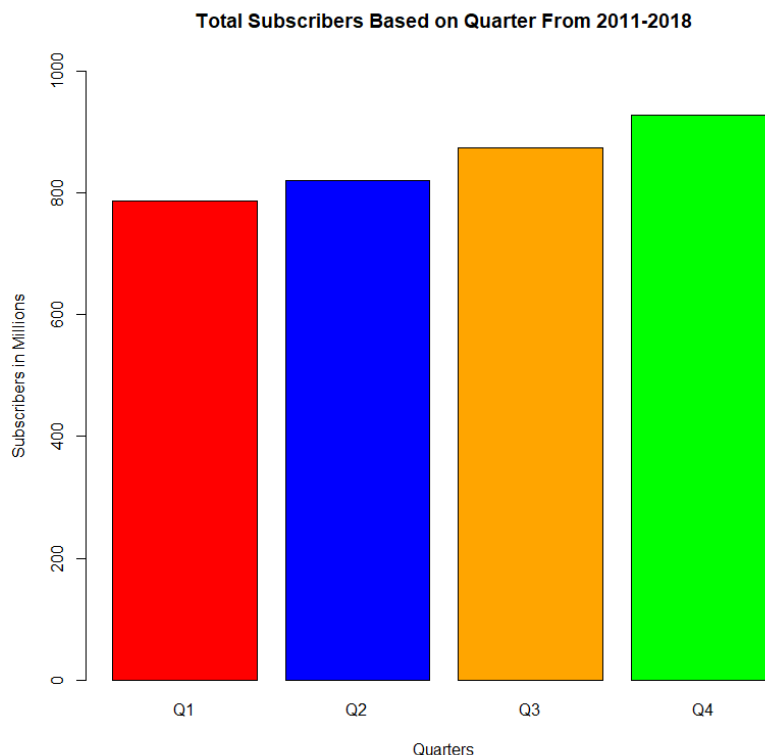
### 3.3 USA Netflix Subscribers From 2011-2018

**USA Netflix Subscribers 2011-2018**



This line graph presents the total USA Netflix Subscribers from 2011 to 2018. Analyzing the graph, there is a positive increase in subscribers over the years until 2018. Netflix subscribers appear to peak at about 203 million subscribers then decrease into 2018. Unfortunately, the USA table stopped in 2018. Is it important to mention that Netflix started their streaming service in 2007 and their sales started to increase substantially in 2011 which is where this table starts. Researching as to why there could be a 35.85 million decrease in subscribers from 2017-2018, the most likely theory is other streaming services heavily competing with Netflix. Services such as YouTube, Hulu and Amazon Prime. Another theory is perhaps there was a newsworthy controversy involving Netflix which led to the streaming service losing subscriptions. Since the data does not include 2018 through 2021, it is difficult to draw conclusions.

### 3.4 Seasonality of Netflix Growth



This bar chart shows a visual representation of the sum of Netflix subscribers from years 2011 and 2018 based on the quarter. The quarters appear to be similar, but there is a 53.58 million subscriber difference in those who joined from Q4 and Q1. Analysing why that might be, quarter 4 represents the months of October, November and December. November and December have major holidays that fall within those months where consumers are making more purchases. Subscription services are known to offer their services for free or discount prices for a limited time to new customers. It is possible that Netflix offered a deal during the part of the year since this is when consumers are purchasing gifts. This same theory could be why quarter 1 subscribers are the lowest since this quarter is coming off the end of the holiday season where consumers tend not to spend a lot of money. It is also possible that consumers who received Netflix subscriptions as a gift simply did not want the service and cancelled it, leading to lower subscribers in those months.

### 3.5 Netflix Growth/Unemployment Correlation

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.35658	1.41067	4.506	6.72e-05
growth_unemploy_cor\$Total	-0.03671	0.22102	-0.166	0.869

According to the linear regression that was designed, the growth in subscribers for Netflix and the unemployment rate have no linear relationship. The R-Squared value is less than .01 which means that less than 1% of our data can be attributed to unemployment fluctuations. With the P-Value being .869, it shows that there is no statistical significance for this correlation. After running a linear regression test, we can conclude that other factors outside of unemployment rates influence the growth rate of Netflix subscriptions.

### 4. Conclusion

For this project, data was collected using businessofapps.com to gather quarterly data from 2011 to 2021 of Netflix subscribers. Also, we used the Bureau of Labor Statistics to gather accurate data on American unemployment rates. Our original goal was to determine if unemployment and streaming service activity had any relationship. Following the collection of data and running a correlation function, we concluded that unemployment and the success of Netflix had no relation. Although this was contrary to our original business problem, we looked deeper into the data collected about Netflix and found some interesting trends. Firstly, it was apparent after generating a histogram displaying quarterly subscribers that quarter 4 tends to have the most subscribers joining the streaming platform. Also, there is a substantial decrease in subscribers during 2018. After some research, we found that Disney announced their release of Disney Plus which caused a decrease in Netflix subscribers. This indicates that Netflix is more influenced by competitors in the streaming industry rather than the economy as a whole. Also, after analyzing our international Netflix histogram, it is clear that the United States had most subscribers in the early 2010's but recently international subscribers have increased substantially.

#### 4.1 Limitations

Our project had a large limitation because we were not able to get monthly data on Netflix. If we were able to have monthly subscriber data, then our monthly unemployment data could show trends that we did not have enough data to find originally. Also, we were limited by our Netflix data because we were not able to get data from their company's initial launch. We had to pick up our analysis from 2011 which does not give us every piece of data that we needed.