# Final Project

Wisdom Uwaifo

2022-07-28

```
library(readr)
library(tidyverse)
library(tidyquant)
library(dplyr)
library(ggplot2)
library(scales)
library(ggrepel)
library(ggtext)
library(patchwork)


Chase_raw <- read_csv("database.csv")
```

#Executive Summary JPMorgan Chase & co is one of the oldest financial institutions in the United States, dating back over 200 years. In 2017, the FDIC announced that Chase had taken over as the bank with the highest number of cash deposits after it had added $96 billion in the previous year. Currently, Chase has total cash deposits worth $2.6 trillion. It has penetrated 100 plus global markets with more than 200 thousand employees. Compared to many other banks in the United States, Chase is undoubtedly one of the most profitable banks in the US, with very loyal customers. One may wonder where these customers come from. That is what my analysis will focus on. In this analysis, I will answer the question, where did Chase Bank customers deposit the most money? I will visualize data to analyze trends in cash deposits over the years. I will show the top 5 states, cities, and bank branches where the most money is deposited. Lastly, I will go further to make comparisons of cash deposits between different years.

#Data background: The dataset used for this analysis includes a cash deposit record for every branch of Chase Bank in the United States, including the branch's name and number, date established as a bank office and (if applicable) acquired by JP Morgan Chase, physical location as street address, city, state, zip, and latitude and longitude coordinates, and the amount deposited at the branch (or the institution, for the bank's main office) between July 1, 2010, to June 30, 2016, in US dollars. The data source is from JPMorgan Chase & Co posted on Kaggle.

#Data cleaning: I did data cleaning for each of the graphs I used for this analysis. The first graph I created is a column graph that shows total cash deposits over the years. To make the graph, I had to rename the variables that were not correctly labeled. I selected just the variables of interest. I used the R pivot_longer() function to combine all the year variables into a single column. I dropped missing values, grouped years, and then summarized the sum of deposits I used to create the total deposits column. Next, I made a column to add dollar signs to the total deposits. For the second graph, I renamed variables and selected just the variables of interest. I filtered the data to determine the top 100 cities and compared their six-year rankings. Hence, I used the distinct function to choose the unique values of cities. I used the pivot longer function to combine the deposits according to years. I grouped by year and then used the ranking function to rank cities. I used the pivot-wider to separate the data to compare them. Then I calculated the ranking difference between the years 2010 and 2016. I dropped missing values and created a condition that ranked cities worse if their ranking decreased by more than 30 and improved if their ranking increased by more than 30. For

the third graph, I only renamed variables again and filtered out the top five cities where the most deposits were made in 2016. Next, I selected the variables of interest and used them to create the graphs. For the final chart, I filtered the top 5 cities where the most deposits were made in 2016. After these cleaning steps, my data was ready for analysis.

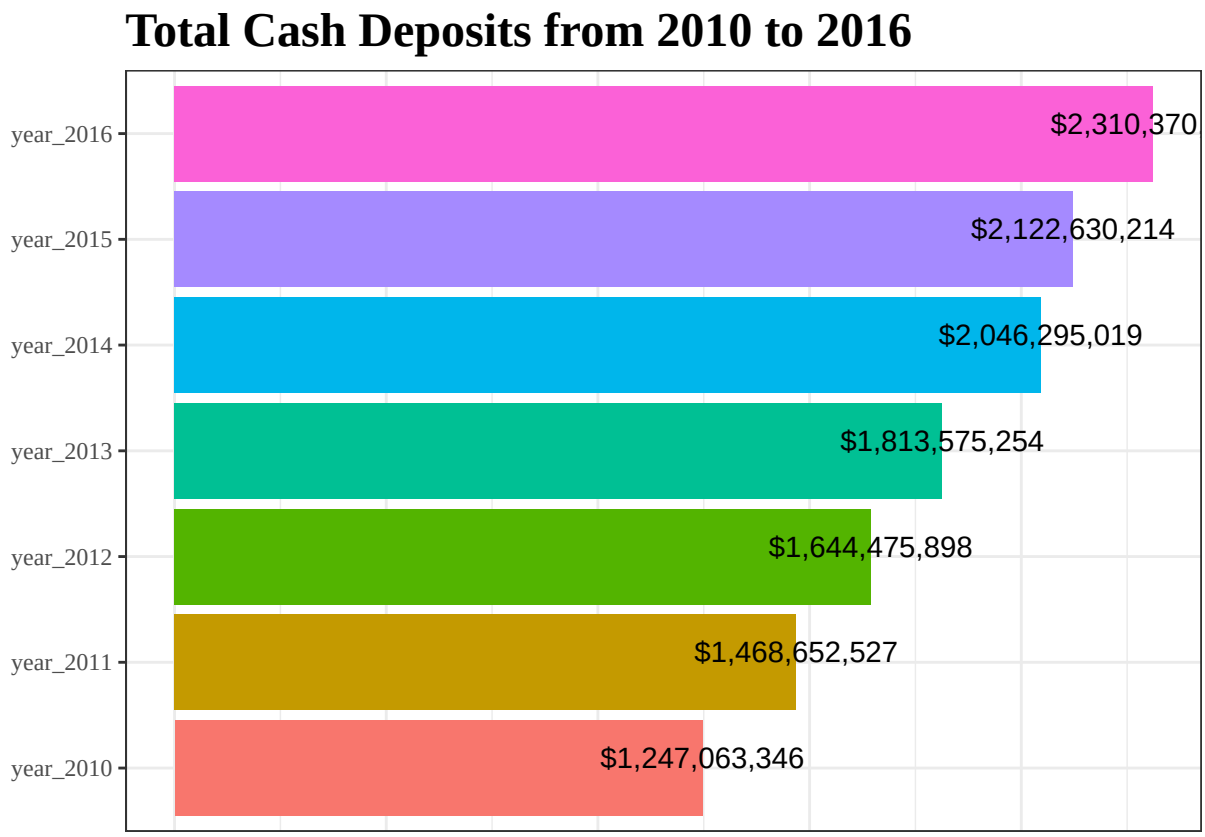
#Individual figures

***Figure 1:***

Figure one is titled "Total Cash Deposits from 2010 to 2016." It is a column chart that shows changes in cash deposits over time. Based on the graph, cash deposits have increased over time from $1.2 trillion in 2010 to $2.3 trillion in 2016. ***Figure 2:*** Figure two compares the six-year rankings of the top 100 cities with the most cash deposits in the United States. The graph compares the ranking between 2010 and 2016 ranking. Cities whose ranking exceeds 30 are colored blue and labeled as improved while the cities whose rankings are decreased by more than 30 are colored red and labeled as worsened. The cities in between are marked as little change and colored grey.
***Figure 3:*** Figure 3 is a lollipop chart that shows the top five branches where the most money was deposited in 2016. ***Final figure:*** I used gravity design to design my project. I added all the graphs I created in R because I wanted to create a storyline. The first show cash deposits over time next shows the top 5 branches of the chase bank where the most deposits were made.

```
Chase_clean <- Chase_raw %>%
  rename(year_2010 = `2010 Deposits`, year_2011 = `2011 Deposits`,
         year_2012 = `2012 Deposits`, year_2013 = `2013 Deposits`,
         year_2014 = `2014 Deposits`, year_2015 = `2015 Deposits`,
         year_2016 = `2016 Deposits`) %>%
  select(year_2010, year_2011, year_2012, year_2013, year_2014,
         year_2015, year_2016) %>%
  pivot_longer(cols = c("year_2010", "year_2011", "year_2012", "year_2013", "year_2014",
         "year_2015", "year_2016",), names_to = "year", values_to = "deposits") %>%
  drop_na(deposits) %>%
  group_by(year) %>%
  summarize(total = sum(deposits)) %>%
  mutate(total2 = dollar(round(total)))
```

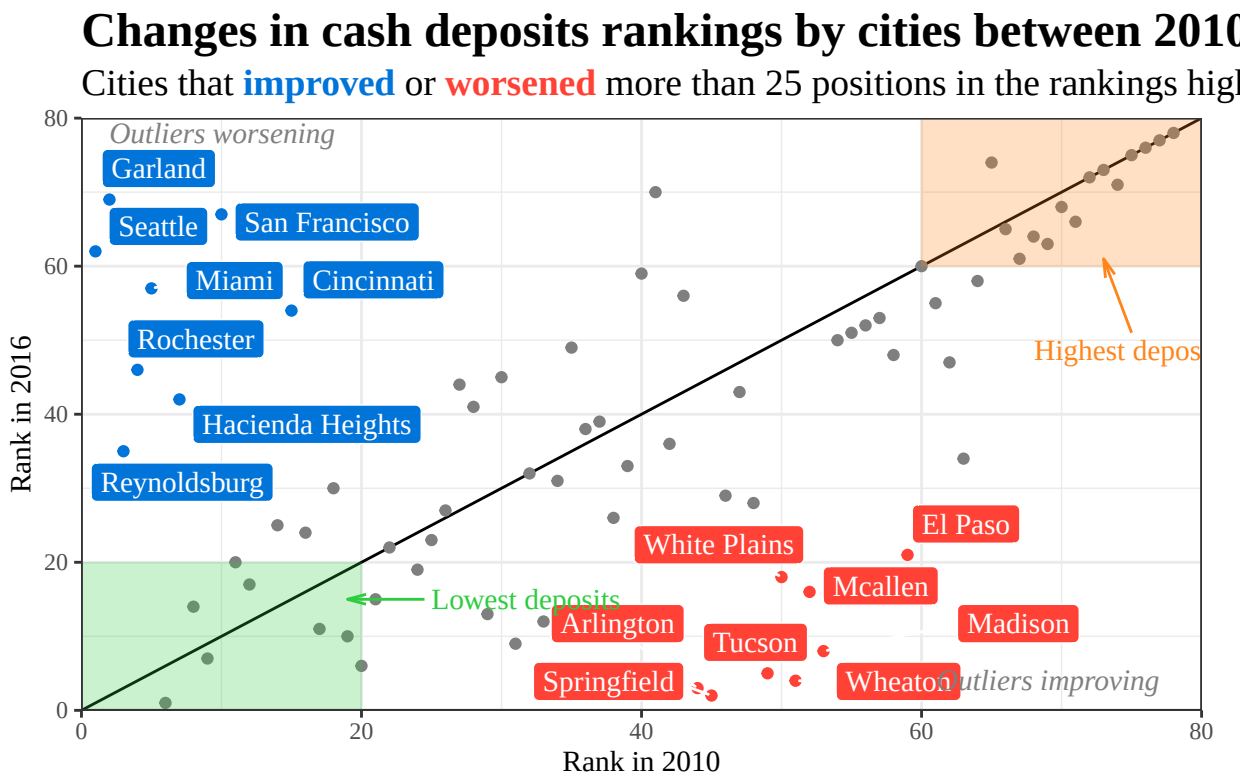
```
Chase_bar <- ggplot(Chase_clean, aes(x = total, y = year, fill = year)) +
  geom_col() +
  geom_text(aes(label = total2), stat = "identity",
            position = "identity", vjust = 0) +
  theme_minimal() +
  guides(fill = "none") +
  labs(title = "Total Cash Deposits from 2010 to 2016") +
  theme_bw(base_family = "serif") +
  theme(plot.title = element_markdown(face = "bold", size = rel(1.6)),
        plot.margin = unit(c(0.5, 1, 0.5, 0.5), units = "lines")) +
    theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank()) +
  scale_size_continuous(labels = comma)
Chase_bar
```




```
Chase_clean2 <- Chase_raw %>%
  rename(year_2010 = `2010 Deposits`, year_2011 = `2011 Deposits`,
         year_2012 = `2012 Deposits`, year_2013 = `2013 Deposits`,
         year_2014 = `2014 Deposits`, year_2015 = `2015 Deposits`,
         year_2016 = `2016 Deposits`) %>%
  select(State, City, year_2010, year_2016) %>%
  drop_na(year_2010, year_2016) %>%
  filter(year_2016 >= 394819) %>%
  distinct(City, .keep_all = TRUE) %>%
  pivot_longer(cols = c("year_2010", "year_2016",), names_to = "year", values_to = "deposits") %>%
  group_by(year) %>%
  mutate(ranking = rank(deposits)) %>%
  ungroup() %>%
  select(State, City, year, ranking) %>%
  pivot_wider(names_from = year, names_prefix = "rank_", values_from = ranking) %>%
  mutate(rank_diff = rank_year_2016 - rank_year_2010) %>%
  drop_na(rank_diff) %>%
  drop_na(rank_diff) %>%
  mutate(big_change = ifelse(abs(rank_diff) >= 30, TRUE, FALSE)) %>%
  mutate(better_big_change = case_when(
    rank_diff <= -30 ~ "Rank worsened",
    rank_diff >= 30 ~ "Rank improved",
    TRUE ~ "Rank changed a little"
  ))
```

```r
Chase_ann <- update_geom_defaults("text", list(family = "serif"))
update_geom_defaults("label", list(family = "serif"))
update_geom_defaults("label_repel", list(family = "serif"))

ggplot(Chase_clean2,
       aes(x = rank_year_2010, y = rank_year_2016)) +
  annotate(geom = "segment", x = 0, xend = 80, y = 0, yend = 80) +
  geom_point(aes(color = better_big_change)) +
  geom_label_repel(data = filter(Chase_clean2, big_change == TRUE),
                   aes(label = City, fill = better_big_change),
                   color = "white") +
  annotate(geom = "text", x = 77, y = 4, label = "Outliers improving",
           fontface = "italic", hjust = 1, color = "grey50") +
    annotate(geom = "text", x = 2, y = 78, label = "Outliers worsening",
           fontface = "italic", hjust = 0, color = "grey50") +
    annotate(geom = "rect", xmin = 0, xmax = 20, ymin = 0, ymax = 20,
           fill = "#2ECC40", alpha = 0.25) +
  annotate(geom = "rect", xmin = 60, xmax = 80, ymin = 60, ymax = 80,
           fill = "#FF851B", alpha = 0.25) +
  annotate(geom = "text", x = 25, y = 15, label = "Lowest deposits",
           hjust = 0, color = "#2ECC40") +
  annotate(geom = "text", x = 75, y = 50, label = "Highest deposits",
           hjust = 0.5, vjust = 1, lineheight = 1, color = "#FF851B") +
  annotate(geom = "segment", x = 75, xend = 73, y = 51, yend = 61, color = "#FF851B",
           arrow = arrow(angle = 15, length = unit(0.5, "lines"))) +
    annotate(geom = "segment", x = 24.5, xend = 19, y = 15, yend = 15, color = "#2ECC40",
           arrow = arrow(angle = 15, length = unit(0.5, "lines"))) +
   scale_color_manual(values = c("grey50", "#0074D9", "#FF4136")) +
  scale_fill_manual(values = c("#0074D9", "#FF4136")) +
  scale_x_continuous(expand = c(0, 0), breaks = seq(0, 80, 20)) +
  scale_y_continuous(expand = c(0, 0), breaks = seq(0, 85, 20)) +
  labs(x = "Rank in 2010", y = "Rank in 2016",
       title = "Changes in cash deposits rankings by cities between 2010 and 2016",
       subtitle = "Cities that <span style='color: #0074D9'>**improved**</span> or <span style='color: #
       caption = "Source: Chase Bank.\nTop 100 Cities with the highest deposits in 2016.") +
  guides(color = "none", fill = "none") +
   theme_bw(base_family = "serif") +
    theme(plot.title = element_markdown(face = "bold", size = rel(1.6)),
        plot.subtitle = element_markdown(size = rel(1.3)),
        plot.margin = unit(c(0.5, 1, 0.5, 0.5), units = "lines"))
```




Source: Chase Bank.
Top 100 Cities with the highest deposits in 2016.

```
Chase_ann
```

```
## NULL
```

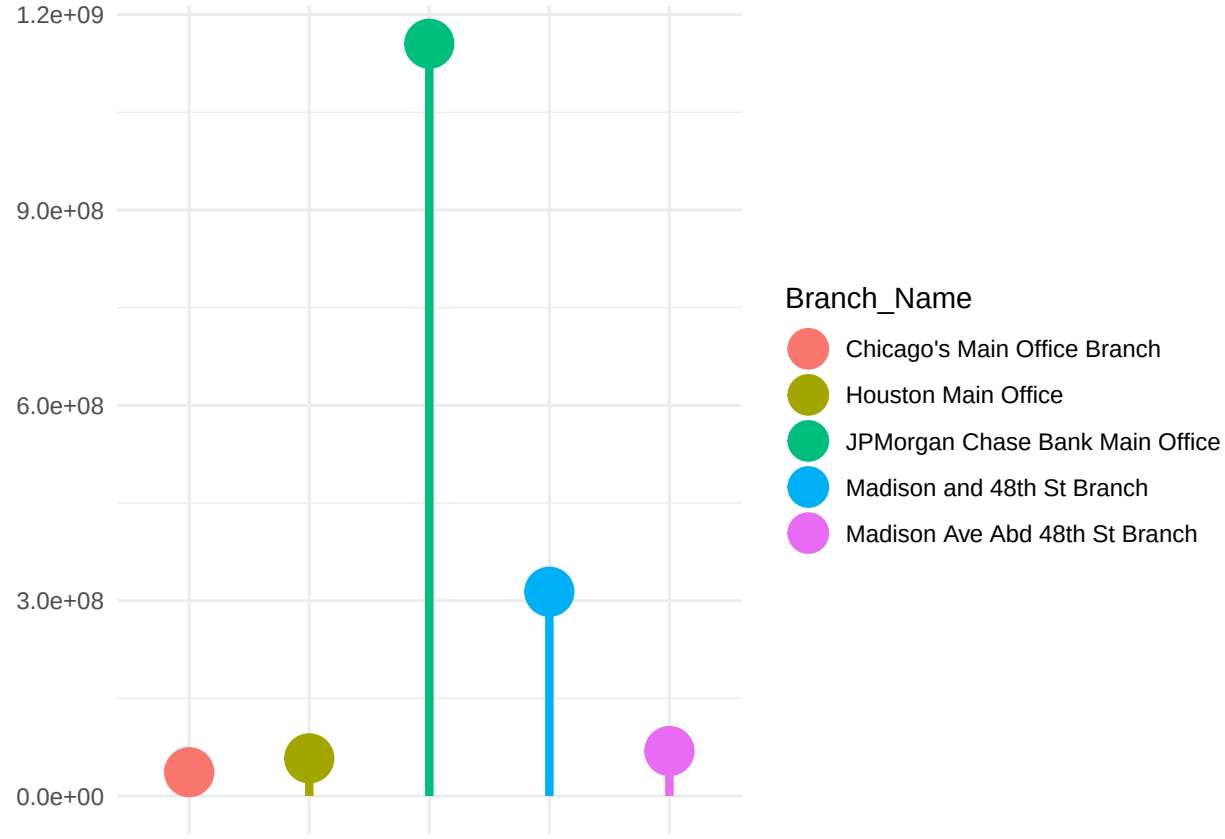
```
Chase_clean3 <- Chase_raw %>%
    rename(year_2010 = `2010 Deposits`, year_2011 = `2011 Deposits`,
         year_2012 = `2012 Deposits`, year_2013 = `2013 Deposits`,
         year_2014 = `2014 Deposits`, year_2015 = `2015 Deposits`,
         year_2016 = `2016 Deposits`, Branch_Name = `Branch Name`) %>%
  filter(year_2016 >= 36539028) %>%
  select(City, Branch_Name, year_2016)
```

```
sum(Chase_clean3$year_2016)
```

```
## [1] 1632276250
```

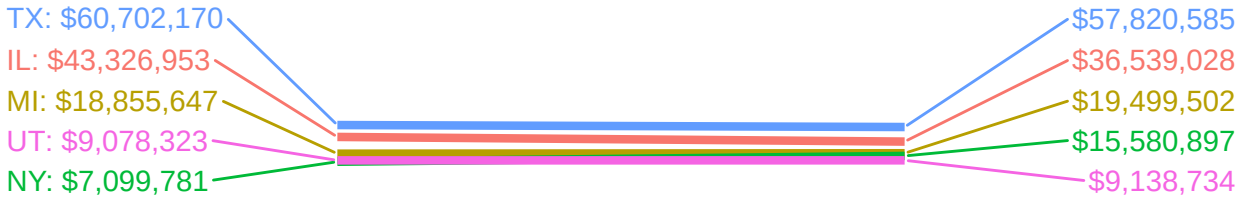
```
Chase_point <- ggplot(data = Chase_clean3,
                      mapping = aes(x = Branch_Name, y = year_2016, color = Branch_Name)) +
  geom_pointrange(aes(ymin = 0, ymax = year_2016), fatten = 5, size = 1.5) +
  theme(strip.background = element_rect(fill = "grey90", color = NA),
        panel.border = element_rect(color = "grey90", fill = NA)) +
  theme_minimal() +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
```

```
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank()) +
  scale_size_continuous(labels = comma)
Chase_point
```




```
Chase_clean4 <- Chase_raw %>%
      rename(year_2010 = `2010 Deposits`, year_2011 = `2011 Deposits`,
         year_2012 = `2012 Deposits`, year_2013 = `2013 Deposits`,
         year_2014 = `2014 Deposits`, year_2015 = `2015 Deposits`,
         year_2016 = `2016 Deposits`, Branch_Name = `Branch Name`) %>%
  select(State, year_2015, year_2016) %>%
   filter(year_2016 >= 8142300) %>%
  distinct(State, .keep_all = TRUE) %>%
  pivot_longer(cols = c("year_2015", "year_2016",), names_to = "year", values_to = "deposits") %>%
  drop_na(deposits) %>%
    mutate(label_first = ifelse(year == "year_2015", paste0(State, ": ", dollar(round(deposits))), NA),
         label_last = ifelse(year == "year_2016", dollar(round(deposits, 0)), NA))
```

```
Chase_line <- ggplot(Chase_clean4, aes(x = year, y = deposits, group = State, color = State)) +
  geom_line(size = 1.5) +
  geom_text_repel(aes(label = label_first), direction = "y", nudge_x = -1, seed = 1234) +
  geom_text_repel(aes(label = label_last), direction = "y", nudge_x = 1, seed = 1234) +
  guides(color = "none") +
  theme_void()
Chase_line
```


```
# Change the dimensions here if you want
ggsave(Chase_bar, filename = "Chase_bar.pdf", device = cairo_pdf,
       width = 7, height = 4, units = "in", bg = "transparent")

ggsave(Chase_ann, filename = "Chase_ann.pdf", device = cairo_pdf,
       width = 7, height = 4, units = "in", bg = "transparent")

ggsave(Chase_point, filename = "Chase_point.pdf", device = cairo_pdf,
       width = 7, height = 4, units = "in", bg = "transparent")

ggsave(Chase_line, filename = "Chase_line.pdf", device = cairo_pdf,
       width = 7, height = 4, units = "in", bg = "transparent")
```