# Data Analysis: 2020 Presidential Election Fraud Analysis via Benford's Law

# Author: Dillan Marroquin

# Date: 05/01/2024 (Originally conducted on 08/11/2021)

# Load libraries

library(tidyverse)

library(scales)

# Set working directory

setwd("C:/Users/di11m/Downloads/Coursera/Data Analysis/RStudio")

# Load election data for Washoe and NV

washoe <- read.csv("Election\_Results\_Washoe\_CLEANED.csv")

nv <- read.csv("Election\_Results\_NV\_CLEANED.csv")

# Define candidate names

candidates <- c("BIDEN, JOSEPH R.", "BLANKENSHIP, DON", "JORGENSEN, JO", "None Of These Candidates", "TRUMP, DONALD J.")

# Function to filter and sum votes per candidate

sum\_votes <- function(data, candidate\_name){

  candidate\_data <- filter(data, Candidate == candidate\_name)

  return(sum(candidate\_data$Votes))

}

# Calculate total votes per candidate: Washoe and NV

biden\_votes\_w <- sum\_votes(washoe, candidates[1])

biden\_votes\_n <- sum\_votes(nv, candidates[1])

blank\_votes\_w <- sum\_votes(washoe, candidates[2])

blank\_votes\_n <- sum\_votes(nv, candidates[2])

jorg\_votes\_w <- sum\_votes(washoe, candidates[3])

jorg\_votes\_n <- sum\_votes(nv, candidates[3])

none\_votes\_w <- sum\_votes(washoe, candidates[4])

none\_votes\_n <- sum\_votes(nv, candidates[4])

trump\_votes\_w <- sum\_votes(washoe, candidates[5])

trump\_votes\_n <- sum\_votes(nv, candidates[5])

# Total votes per candidate: Washoe and NV

total\_votes\_w <- sum(biden\_votes\_w, blank\_votes\_w, jorg\_votes\_w, none\_votes\_w, trump\_votes\_w)

total\_votes\_n <- sum(biden\_votes\_n, blank\_votes\_n, jorg\_votes\_n, none\_votes\_n, trump\_votes\_n)

# Total votes by candidate data frame: Washoe and NV

votes\_by\_candidate\_w <- data.frame(

  Candidate = candidates,

  Votes = c(biden\_votes\_w, blank\_votes\_w, jorg\_votes\_w, none\_votes\_w, trump\_votes\_w)

)

votes\_by\_candidate\_n <- data.frame(

  Candidate = candidates,

  Votes = c(biden\_votes\_n, blank\_votes\_n, jorg\_votes\_n, none\_votes\_n, trump\_votes\_n)

)

# Percentage of votes per candidate: Washoe and NV

percents\_w <- votes\_by\_candidate\_w$Votes / total\_votes\_w

percents\_n <- votes\_by\_candidate\_n$Votes / total\_votes\_n

# Percentage of votes per candidate data frame: Washoe and NV

percent\_total\_w <- data.frame(

  Candidate = candidates,

  Percent = percents\_w \* 100

)

percent\_total\_n <- data.frame(

  Candidate = candidates,

  Percent = percents\_n \* 100

)

# Graph generator function

generate\_graph <- function(candidate\_names, votes, title){

  # Data frame for graph

  graph\_data <- data.frame(

    Candidate = candidate\_names,

    Votes = votes

  )

  # Generate graph

  graph <- ggplot(graph\_data, aes(x = Candidate, y = Votes, fill = Candidate)) +

    geom\_col() +

    theme(legend.position = "none") +

    scale\_y\_continuous(breaks = scales::pretty\_breaks(), labels = scales::comma\_format()) +

    scale\_x\_discrete(guide = guide\_axis(n.dodge = 2)) +

    scale\_fill\_manual(values = c(

      "BIDEN, JOSEPH R." = "blue",

      "BLANKENSHIP, DON" = "purple",

      "JORGENSEN, JO" = "yellow",

      "None Of These Candidates" = "black",

      "TRUMP, DONALD J." = "red")) +

    labs(

      title = title,

      x = "Candidate",

      y = "Votes"

    )

  return(graph)

}

# Graph of election results: Washoe and NV

election\_graph\_w <- generate\_graph(

  candidate\_names = candidates,

  votes = c(biden\_votes\_w, blank\_votes\_w, jorg\_votes\_w, none\_votes\_w, trump\_votes\_w),

  title = "2020 Presidential Election Results: Washoe County"

  )

election\_graph\_n <- generate\_graph(

  candidate\_names = candidates,

  votes = c(biden\_votes\_n, blank\_votes\_n, jorg\_votes\_n, none\_votes\_n, trump\_votes\_n),

  title = "2020 Presidential Election Results: Nevada"

)

# Benford's Law distribution

bens\_law <- function(d) log10(1 + 1 / d)  #Formula

digits <- 1:9

bens\_data <- data.frame(Digits = digits, Percentage = bens\_law(digits))

bens\_graph <-

  ggplot(bens\_data,

         aes(x = Digits, y = Percentage)) +

  geom\_line(color = "red") +

  geom\_point(color = "red", size = 3) +

  scale\_x\_continuous(n.breaks = 9) +

  scale\_y\_continuous(n.breaks = 8) +

  labs(title = "Benford's Law",

       x = "Leading Digit",

       y = "Proportion of Occurrence")

# Leading Digits: Washoe

ld\_occur\_w <- sapply(1:9, function(x) sum(washoe$LeadDigit == x))

ld\_percent\_w <- ld\_occur\_w / sum(ld\_occur\_w)

ld\_data\_w <- data.frame(Digits = digits, Percentage = ld\_percent\_w)

ld\_graph\_w <-

  ggplot(ld\_data\_w,

         aes(x = Digits, y = Percentage)) +

  geom\_line(color = "blue") +

  geom\_point(color = "blue", size = 3) +

  scale\_x\_continuous(n.breaks = 9) +

  scale\_y\_continuous(n.breaks = 8) +

  labs(title = "Leading Digit of Total Votes\nin Washoe Presidential Election",

       x = "Leading Digit",

       y = "Proportion of Occurrence")

# Leading Digits: NV

ld\_occur\_n <- sapply(1:9, function(x) sum(washoe$LeadDigit == x))

ld\_percent\_n <- ld\_occur\_n / sum(ld\_occur\_n)

ld\_data\_n <- data.frame(Digits = digits, Percentage = ld\_percent\_w)

ld\_graph\_n <-

  ggplot(ld\_data\_n,

         aes(x = Digits, y = Percentage)) +

  geom\_line(color = "orange") +

  geom\_point(color = "orange", size = 3) +

  scale\_x\_continuous(n.breaks = 9) +

  scale\_y\_continuous(n.breaks = 8) +

  labs(title = "Leading Digit of Total Votes\nin Nevada Presidential Election",

       x = "Leading Digit",

       y = "Proportion of Occurrence")

# Benford's Law comparison: Washoe

table\_wb <- data.frame(Digits = digits,

                      Percentage = c(bens\_law(digits), ld\_percent\_w),

                      Focus = c(rep("Benford", 9), rep("Washoe", 9)))

graph\_wb <-

  ggplot(table\_wb,

         aes(x = Digits, y = Percentage, group = Focus)) +

  geom\_line(aes(color = Focus)) +

  geom\_point(aes(color = Focus)) +

  scale\_x\_continuous(n.breaks = 9) +

  scale\_y\_continuous(n.breaks = 8) +

  scale\_color\_manual(values = c("red", "blue")) +

  labs(title = "Leading Digits of Votes in Washoe v. Benford",

       x = "Leading Digit",

       y = "Proportion of Occurrence")

# Benford's Law comparison: NV

table\_nb <- data.frame(Digits = digits,

                      Percentage = c(bens\_law(digits), ld\_percent),

                      Focus = c(rep("Benford", 9), rep("Nevada", 9)))

graph\_nb <-

  ggplot(table\_nb,

         aes(x = Digits, y = Percentage, group = Focus)) +

  geom\_line(aes(color = Focus)) +

  geom\_point(aes(color = Focus)) +

  scale\_x\_continuous(n.breaks = 9) +

  scale\_y\_continuous(n.breaks = 8) +

  scale\_color\_manual(values=c("red", "orange")) +

  labs(title = "Leading Digits of Votes in Nevada v. Benford",

       x = "Leading Digit",

       y = "Proportion of Occurrence")

# Benford's Law comparison: NV, Washoe, Benford

table\_all <- data.frame(Digits = digits,

                       Percentage = c(bens\_law(digits), ld\_percent\_w, ld\_percent\_n),

                       Focus = c(rep("Benford", 9), rep("Washoe", 9), rep("Nevada", 9)))

graph\_all <-

  ggplot(table\_all,

         aes(x = Digits, y = Percentage, group = Focus)) +

  geom\_line(aes(color = Focus)) +

  geom\_point(aes(color = Focus)) +

  scale\_x\_continuous(n.breaks = 9) +

  scale\_y\_continuous(n.breaks = 8) +

  scale\_color\_manual(values = c("red", "orange", "blue")) +

  labs(title = "Washoe and Nevada v. Benford",

       x = "Leading Digit",

       y = "Proportion of Occurrence")