

Logan_final

Andrew vanderWilden

5/29/2020

```
library(tidyverse)
library(gt)
library(glue)
library(scales)
library(lubridate)
```

```
depart <- read.csv("Departures.csv", skip = 7)
airline <- read.csv("Airline_codes.csv")
```

```
# Read in Airport Data
airports <- read.csv("airports.dat",
                    colClasses = "character",
                    col.names = c("airport.id",
                                   "name",
                                   "city",
                                   "country",
                                   "iata",
                                   "icao",
                                   "latitude",
                                   "longitude",
                                   "altitude",
                                   "time.zone",
                                   "dst",
                                   "tz",
                                   "type",
                                   "source"),
                    header = FALSE)
```

```
airports <- airports[,c(2,3,4,5,7,8,9)]
```

```
# Average Delays by Destination
```

```
delays <- depart %>%
  group_by(Destination.Airport) %>%
  summarise(N_Flights = n(),
            Avg_delay = mean(Departure.delay..Minutes.))

delays <- delays %>%
  left_join(airports, by = c("Destination.Airport" = "iata")) %>%
```

```

  filter(country == "United States")

delays$b_long <- -71.0096
delays$b_lat <- 42.3656

delays$latitude <- as.double(delays$latitude)
delays$longitude <- as.double(delays$longitude)

map_dat <- map_data("state") %>%
  fortify() %>%
  as_tibble()

delmap <- map_dat %>%
  ggplot(aes(long,lat)) +
  geom_map(map = map_dat, aes(map_id = region), fill = "#03731d", alpha = 0.6) +
  borders("state", colour = "black") +
  geom_point(
    data = delays,
    aes(x = as.numeric(longitude),
        y = as.numeric(latitude),
        size = N_Flights,
        color = Avg_delay),
    alpha = 0.8) +
  scale_color_gradient(low = "#f6fa05", high = "#ba0802") +
  geom_curve(data = delays,
    aes(x = b_long,
        xend = longitude,
        y = b_lat,
        yend = latitude,
        color = Avg_delay,
        size = 0.005),
    curvature = 0.2,
    alpha = 0.25,
    show.legend = FALSE,
    inherit.aes = FALSE) +
  theme(
    legend.position = "bottom",
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 11),
    axis.title = element_blank(),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
    panel.background = element_rect(fill = "#04bdde")
  ) +
  guides(
    color = guide_colorbar(title = "Average Delay in Minutes",
      direction = "horizontal",
      title.position = "top",
      title.hjust = 0.2,
      title.vjust = 0.85),
    alpha = guide_none(),

```

```

size = guide_legend(title = "Total Flights",
                    title.position = "top",
                    title.hjust = 0.5,
                    title.vjust = 0.85)

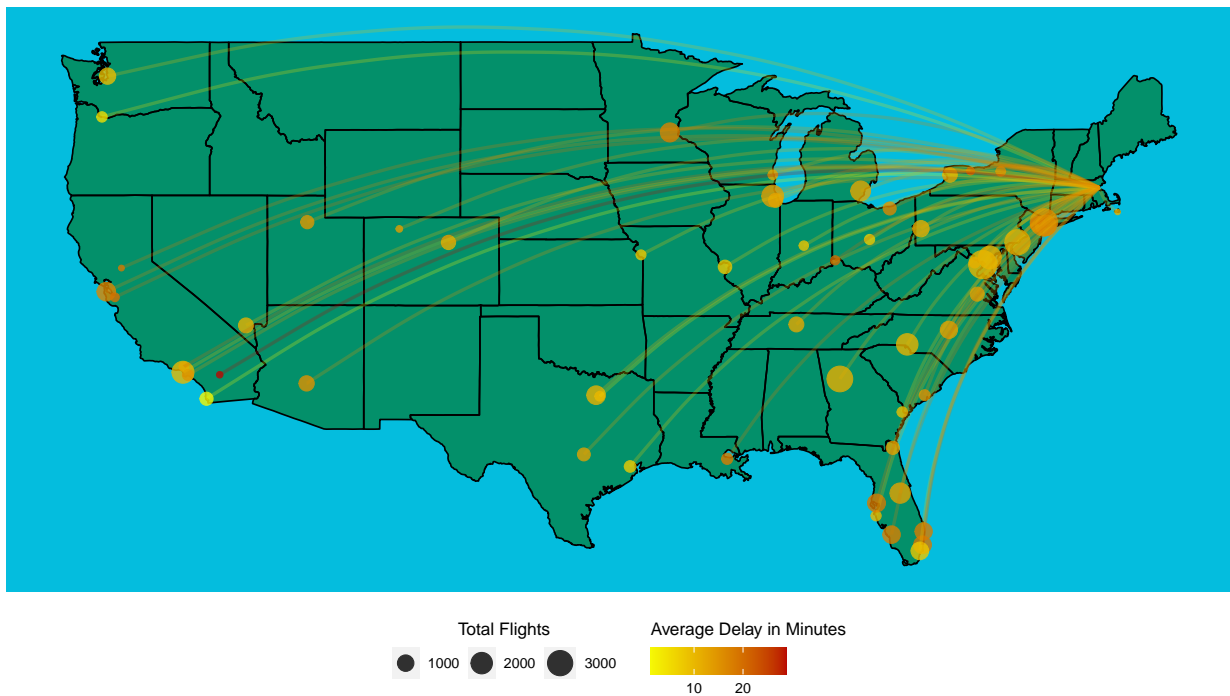
) +
scale_size(range = c(1,8)) +
labs(title = "Average Flight Delays from Logan Airport to Destinations Around the US",
     subtitle = "October 2018 - March 2019",
     caption = "Data from United States Department of Transportation")

```

delmap

Average Flight Delays from Logan Airport to Destinations Around the US

October 2018 – March 2019



Data from United States Department of Transportation

```

del_inf <- delays %>% top_n(10, Avg_delay) %>% arrange(-Avg_delay)

del_inf <- del_inf[,c(1,4,5,2,3)]

# Could Format Delay as MM:SS ie 20:35 but then the color scale does not work

#del_inf$Avg_delay <- ifelse(round((del_inf$Avg_delay-floor(del_inf$Avg_delay))*60) >= 10,
#                             # paste(floor(del_inf$Avg_delay),
#                             #       round((del_inf$Avg_delay-floor(del_inf$Avg_delay))*60),
#                             #       sep = ":"),
#                             # paste(floor(del_inf$Avg_delay),
#                             #       round((del_inf$Avg_delay-floor(del_inf$Avg_delay))*60),
#                             #       sep = ":0"))

```

```

colnames(del_inf) <- c("IATA Code", "Airport", "City", "Flights", "Average Delay")

del_cols <- c("#f6fa05", "#ba0802")
delay_scale <- col_numeric(del_cols, domain = c(min(del_inf$`Average Delay`),
                                                max(del_inf$`Average Delay`)))

delay_gt <- del_inf %>%
  gt() %>%
  tab_source_note(md("**SOURCE**: United States Department of Transportation")) %>%
  tab_header(title = md("**Longest Average Delays by Destination from Logan Airport**"),
             subtitle = "Domestic Flights | October 2018 - March 2019") %>%
  fmt_number("Flights", decimals = 0) %>%
  fmt_number("Average Delay", decimals = 2) %>%
  tab_footnote(footnote = "Measured in Minutes",
               locations = cells_column_labels(
                 columns = vars(`Average Delay`)
               )) %>%
  opt_footnote_marks(marks = "standard") %>%
  tab_style(
    style = cell_text(
      size = px(13),
      font = "arial",
      transform = "uppercase",
      weight = "bold"
    ),
    locations = cells_column_labels(everything())
  ) %>%
  tab_style(
    style = cell_text(
      indent = px(40)
    ),
    locations = cells_body(
      vars(`Flights`)
    )
  ) %>%
  tab_style(
    style = cell_text(
      indent = px(40)
    ),
    locations = cells_column_labels(
      vars(`Flights`)
    )
  ) %>%
  tab_style(
    style = cell_text(
      align = "center"
    ),
    locations = cells_body(
      vars(`Average Delay`)
    )
  ) %>%
  tab_style(

```

```

    style = cell_text(
      indent = px(25)
    ),
    locations = cells_column_labels(
      vars(`City`)
    )
  ) %>%
  tab_style(
    style = cell_text(
      indent = px(25)
    ),
    locations = cells_body(
      vars(`City`)
    )
  ) %>%
  tab_style(
    style = cell_text(
      indent = px(15)
    ),
    locations = cells_column_labels(
      vars(`Airport`)
    )
  ) %>%
  tab_style(
    style = cell_text(
      indent = px(15)
    ),
    locations = cells_body(
      vars(`Airport`)
    )
  ) %>%
  data_color(
    columns = vars(`Average Delay`),
    colors = delay_scale
  ) %>%
  tab_style(
    style = cell_borders(
      side = "bottom", color = "black", weight = px(3)
    ),
    locations = cells_column_labels(everything())
  )

```

delay_gt

Longest Average Delays by Destination from Logan Airport

Domestic Flights | October 2018 - March 2019

IATA Code	Airport	City	Flights	Average Delay *
PSP	Palm Springs International Airport	Palm Springs	27	28.93
EWR	Newark Liberty International Airport	Newark	847	23.32
ROC	Greater Rochester International Airport	Rochester	83	22.13

BUR	Bob Hope Airport	Burbank	173	20.61
CVG	Cincinnati Northern Kentucky International Airport	Cincinnati	150	20.16
SJC	Norman Y. Mineta San Jose International Airport	San Jose	127	19.35
FLL	Fort Lauderdale Hollywood International Airport	Fort Lauderdale	1,220	18.62
TPA	Tampa International Airport	Tampa	1,258	18.49
SMF	Sacramento International Airport	Sacramento	11	18.45
PBI	Palm Beach International Airport	West Palm Beach	1,145	18.04

*Measured in Minutes

SOURCE: United States Department of Transportation

```
raincloud_theme <- theme(
  text = element_text(size = 10),
  axis.title.x = element_text(size = 16),
  axis.title.y = element_text(size = 16),
  axis.text = element_text(size = 14),
  axis.text.x = element_text(angle = 45, vjust = 0.5),
  legend.title = element_text(size = 16),
  legend.text = element_text(size = 16),
  legend.position = "right",
  plot.title = element_text(lineheight = 0.8, face = "bold", size = 16),
  panel.border = element_blank(),
  panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(),
  axis.line.x = element_line(colour = "black", size = 0.5, linetype = "solid"),
  axis.line.y = element_line(colour = "black", size = 0.5, linetype = "solid")
)
```

```
c_dat1 <- depart[,c("Carrier.Code", "Departure.delay..Minutes.")]
colnames(c_dat1) <- c("Airline", "Delay")
```

```
source("https://gist.githubusercontent.com/benmarwick/2a1bb0133ff568cbe28d/raw/fb53bd97121f7f9ce947837e")
```

```
rcp <- c_dat1 %>%
  ggplot(aes(Airline, Delay, fill = Airline)) +
  geom_flat_violin(position = position_nudge(x = .2, y = 0), alpha = 0.8) +
  geom_point(aes(y = Delay, color = Airline),
    position = position_jitter(width = 0.15), size = 0.5, alpha = 0.8) +
  geom_boxplot(width = 0.1, outlier.shape = NA, alpha = 0.5) +
  guides(fill = FALSE) +
  guides(color = FALSE) +
  coord_flip() +
  theme_bw() +
  ggtitle("Raincloud Plot of Delays by Airline") +
  raincloud_theme +
  ylim(-5,200)

rcp
```

Raincloud Plot of Delays by Airline

