*The University of Texas at Austin*

*Zhaowen Fan*

*Rafael Ignacio Gonzalez Chong*

*Assignment 7*

GSI Intro to Big Data and Data Mining

Table of Contents

[Create at least two of the following Visualization Ideas or create your own visualization idea using ggplot2 package. 2](#_Toc204324930)

[Appendices (Code) 3](#_Toc204324931)

# **Create at least two of the following Visualization Ideas or create your own visualization idea using ggplot2 package.**

A graph of a number of flights per day

AI-generated content may be incorrect.

Fig 1. Number of flights per day of the week

A graph showing the departure of a flight

AI-generated content may be incorrect.

Fig. 2 Average departure delay per day of the week

# **Appendices (Code)**

#ASSIGNMENT 7

#GSI Intro to Big Data and Data Mining

#Zhaowen Fan

#Rafael Ignacio Gonzalez Chong

library(dplyr)

library(ggplot2)

flights.file <- "/Users/rafaelgonzalez/Desktop/assignment7/flights-small.csv"

flights <- read.csv(flights.file, stringsAsFactors = FALSE)

#Create at least two of the following Visualization Ideas,

#or create your own visualization idea using ggplot2 package.

#Number of flights per day of the week

flights %>%

count(DAY\_OF\_WEEK) %>%

ggplot(aes(

x = factor(DAY\_OF\_WEEK, levels = 1:7, labels = c("Mon", "Tue", "Wed", "Thu", "Fri", "Sat", "Sun")),

y = n,

fill = n

)) +

geom\_col() +

scale\_fill\_gradient(low = "lightgreen", high = "darkgreen") +

labs(

title = "Number of Flights per Day",

x = "Day",

y = "Flights"

)

#Average departure delay per day of the week

flights %>%

filter(CANCELLED == 0, !is.na(DEPARTURE\_DELAY)) %>%

group\_by(DAY\_OF\_WEEK) %>%

summarise(avg\_delay = mean(DEPARTURE\_DELAY)) %>%

ggplot(aes(factor(DAY\_OF\_WEEK, levels=1:7, labels=c("Mon","Tue","Wed","Thu","Fri","Sat","Sun")), avg\_delay, fill=avg\_delay)) +

geom\_col() +

labs(title="Avg Departure Delay by Day", x="Day", y="Avg Delay in min")