Midterm

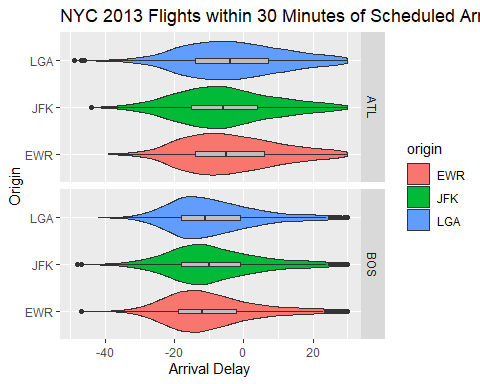
Vana Venjimuri

10/12/2019

# Question 1

flights %>%  
 filter(dest %in% c("ATL", "BOS")) %>%  
 filter(origin %in% c("LGA", "JFK", "EWR")) %>%  
 filter(arr\_delay <= 30) %>%  
 ggplot(aes(  
 x = origin ,  
 y = arr\_delay ,  
 fill = origin,  
 legend = FALSE  
 )) +  
 geom\_violin(legend = FALSE) +  
 geom\_boxplot(width = 0.1 , fill = "grey") +  
 coord\_flip() +  
 facet\_grid(rows = vars(dest)) +  
 labs(title = "NYC 2013 Flights within 30 Minutes of Scheduled Arrival",  
 x = "Origin",  
 y = "Arrival Delay")

## Warning: Ignoring unknown parameters: legend



# Question 2

not\_cancelled <- flights %>%  
filter(!is.na(dep\_delay),!is.na(arr\_delay))  
  
not\_cancelled %>%  
filter(origin %in% c("LGA", "JFK", "EWR")) %>%  
group\_by(origin) %>%  
filter(arr\_delay < 0) %>%  
summarize(flight = n(),  
on\_time\_percentage = mean(arr\_delay <= 0, na.rm = TRUE) \* 100) %>%  
arrange(desc(on\_time\_percentage)) %>%  
ggplot(aes(  
x = origin,  
y = on\_time\_percentage ,  
fill = origin,  
legend = FALSE  
)) +  
geom\_bar(stat = "identity", legend = FALSE) +  
labs(title = "Bar Chart (Pareto) of Percentage on Time",  
y = "Percentage")

## Warning: Ignoring unknown parameters: legend

