Homework1

Aishwarya

10/4/2021

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(nycflights13)

ewr\_flights <- flights %>%   
 filter(origin == "EWR" & (carrier == "EV" | carrier == "UA")) %>%  
 mutate(late = (dep\_delay > 0)) %>%  
 mutate(very\_late = (dep\_delay > 30))  
ewr\_flights

## # A tibble: 90,026 × 21  
## year month day dep\_time sched\_dep\_time dep\_delay arr\_time sched\_arr\_time  
## <int> <int> <int> <int> <int> <dbl> <int> <int>  
## 1 2013 1 1 517 515 2 830 819  
## 2 2013 1 1 554 558 -4 740 728  
## 3 2013 1 1 558 600 -2 923 937  
## 4 2013 1 1 559 600 -1 854 902  
## 5 2013 1 1 607 607 0 858 915  
## 6 2013 1 1 624 630 -6 909 840  
## 7 2013 1 1 628 630 -2 1016 947  
## 8 2013 1 1 632 608 24 740 728  
## 9 2013 1 1 643 646 -3 922 940  
## 10 2013 1 1 644 636 8 931 940  
## # … with 90,016 more rows, and 13 more variables: arr\_delay <dbl>,  
## # carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,  
## # air\_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time\_hour <dttm>,  
## # late <lgl>, very\_late <lgl>

total\_mean\_proportion <- ewr\_flights %>%   
 group\_by(carrier) %>%   
 summarize(proportion\_late\_mean = mean(late, na.rm = TRUE),proportion\_verylate\_mean = mean(very\_late, na.rm = TRUE))  
total\_mean\_proportion

## # A tibble: 2 × 3  
## carrier proportion\_late\_mean proportion\_verylate\_mean  
## <chr> <dbl> <dbl>  
## 1 EV 0.462 0.230  
## 2 UA 0.498 0.135