HPAM 7660 Data Assignment 3

Rebecca Letsinger

February 29, 2024

Question # 1 {r} install.packages("dplyr") install.packages("knitr") install.packages("nycflights13") {r} library(dplyr) library(knitr) library(nycflights13)

Question # 2 {r} flights <- flights %>% mutate(avg_speed = distance / (air_time / 60))

Question # 3 {r} carrier_avg_speed <- flights %>% group_by(carrier) %>% summarize(avg_speed = mean(distance / (air_time / 60), na.rm = TRUE))

{r} kable(carrier_avg_speed, caption = "Carrier-specific average air speeds")

Question # 4 {r} carrier_summary <- flights %>% group_by(carrier) %>% summarize(avg_speed = mean(distance / (air_time / 60), na.rm = TRUE), sd_speed = sd(distance / (air_time / 60), na.rm = TRUE), min_speed = min(distance / (air_time / 60), na.rm = TRUE), max_speed = max(distance / (air_time / 60), na.rm = TRUE), num_obs = n())

{r} kable(carrier_summary, caption = "Summary statistics of carrier-specific average air speeds")

Question # 5 {r} carrier_summary <- flights %>% group_by(carrier) %>% summarize(avg_speed = mean(distance / (air_time / 60), na.rm = TRUE), sd_speed = sd(distance / (air_time / 60), na.rm = TRUE), min_speed = min(distance / (air_time / 60), na.rm = TRUE), max_speed = max(distance / (air_time / 60), na.rm = TRUE), num_obs = n()) %>% arrange(desc(avg_speed))

{r} kable(carrier_summary, caption = "Summary statistics of carrier-specific average air speeds (sorted by average air speed)")

Question # 6 {r} carrier_summary <- flights %>% group_by(carrier) %>% summarize(avg_speed = mean(distance / (air_time / 60), na.rm = TRUE), sd_speed = sd(distance / (air_time / 60), na.rm = TRUE), min_speed = min(distance / (air_time / 60), na.rm = TRUE), max_speed = max(distance / (air_time / 60), na.rm = TRUE), num_obs = n()) %>% arrange(desc(avg_speed))

{r} carrier_summary <- left_join(carrier_summary, airlines, by = c("carrier" = "carrier")) %>% select(-carrier) %>% # Remove the carrier column rename(airline = name)

{r} kable(carrier_summary, caption = "Summary statistics of carrier-specific average air speeds with carrier names")

Question # 7: {r} joined_data <- flights %>% left_join(weather, by = c("origin", "year", "month", "day", "hour"))

{r} carrier_summary <- joined_data %>% group_by(carrier) %>% summarize(avg_speed = mean(distance / (air_time / 60), na.rm = TRUE), avg_humidity = mean(humid, na.rm = TRUE), sd_speed = sd(distance / (air_time / 60), na.rm = TRUE), min_speed = min(distance / (air_time / 60), na.rm = TRUE), max_speed = max(distance / (air_time / 60), na.rm = TRUE), num_obs = n()) %>% arrange(desc(avg_speed))

{r} kable(carrier_summary, caption = "Summary statistics of carrier-specific average air speeds with average humidity")

Question # 8 {r} carrier_summary <- carrier_summary %>% select(airline, avg_speed, sd_speed, min_speed, max_speed, avg_humidity)

{r} kable(carrier_summary, caption = "Summary statistics of carrier-specific average air speeds with average humidity")