MATH113/CAIS105: Intro to Data Science

Fall 2023

Homework 03

Homework is DUE before class on the day indicated on the course schedule.

Learning Objectives:

- Practice data wrangling and analysis of 1 table in R

Part 1

Complete this assignment in an R Markdown file.

Answer the following questions using the `nycflights13` package:

1. What was the daily average number of flights leaving each of the three NYC airports in 2013?
2. For each carrier, compute the number of total flights, the average departure delay, the number of unique destinations serviced, and the number of unique planes used.
3. Plot the distribution of average daily delay time across the entire year for each of the three airports.

When you answer these questions, be sure to include your code *and* a written answer in your R Markdown file. For example, if I were answering the question: "What were the most popular baby names in the 1990s", my R Markdown report would look something like:

```
babynames %>%
  filter(year >= 1990 & year < 2000) %>%
  group_by(name) %>%
  summarize(num_births = sum(n)) %>%
  arrange(desc(num_births))
```

```
## # A tibble: 45,928 x 2
##    name        num_births
##    <chr>            <int>
##  1 Michael         464249
##  2 Christopher     361251
##  3 Matthew         352341
##  4 Joshua          330046
##  5 Jessica         303854
##  6 Ashley          303125
##  7 Jacob           298926
##  8 Nicholas        275906
##  9 Andrew          273515
## 10 Daniel          273347
## # … with 45,918 more rows
```

The most popular baby names from the 1990s were Michael, Christopher, and Matthew.

Optional Challenge: Plot the average arrival delay time as a function of the distance flown *to the nearest 100 miles* for each of the three airports.

Submission

Knit your R Markdown file to a PDF and submit through PLATO.