

# Lab 09: Strings

Owen G. Ward

```
library(tidyverse)
library(nycflights13)
```

In this lab we will do some basic string manipulation using the `flights` data.

1.

The `tailnum` variable is a unique ID for a plane. We want to see how many different possible lengths this ID can have, in this dataset, and if it varies across carriers.

To do this

- Create a new variable `tail_len` which gives the number of characters in the tail number for each flight.
  - Use `group_by` to get the `mean` and `sd` of `tail_len` for each carrier. You will need to deal with NA values. How many carriers have planes with tail numbers shorter than 6?
2. Create a new variable `carr_id` in this dataset which combines the carrier code and the tail number, with the separator `_`.
3. In the `planes` dataset, Airbus planes are given in the `model` variable by an ID that starts with `A` followed by numbers. The first 3 numbers after `A` are most important. Write a regular expression to extract these codes (`A` and the 3 numbers directly after it) from the `model` column. Store this in a vector, keeping only the non missing values.