Relational Data Lab

2019-03-08

We'll work more with the nycflights13 dataset here.

1. Load the nycflights13 data frames into R.
2. Explore if the average arrival delay is associated with the time zone of the destination.
3. Calculate the average longitude and latitude of the three NCY airports. We will use this as the location of NYC.
4. Create a function that takes two vectors (`lon` and `lat`) and two scalars (`nyclon` and `nyclat`) as input. It then uses the `distm()` function in the geosphere R package to calculate the geodesic distance between each each element in `lon` and `lat` and the location defined by `nyclon` and `nyclat`. For example, in my implementation:

```
lat    <- c(41.1, 32.5, 42.0)
lon    <- c(-80.6, -85.7, -88.1)
nyclon <- -73.9
nyclat <- 40.8
dist_nyc(lon = lon, lat = lat, nyclon = nyclon, nyclat = nyclat)
```

```
## [1]  564974 1398237 1193526
```

5. Use the function you created in part 4 and the NYC location you calculated in part 3 to find the geodesic distance from NYC of every airport. Explore if average arrival delay is associated with distance of the destination from NYC.
6. For each plane calculate the proportion of times that it takes off in fair weather and see if that proportion is associated with the age of the plane. We'll define "fair weather" to mean no precipitation, wind-speeds of under 20 mph, and a visibility of at least 10 miles.