

# Exercise set 1

1. Imagine you wanted to draw (approximately) the route each plane flies from its origin to its destination. What variables would you need? What tables would you need to combine?
2. We know that some days of the year are “special”, and fewer people than usual fly on them. How might you represent that data as a data frame? What would be the primary keys of that table? How would it connect to the existing tables?

# Sometimes tables don't have explicit primary keys

flights

```
#> # A tibble: 336,776 x 19
```

```
#>   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     533           529           4       850           830
#> 3  2013     1     1     542           540           2       923           850
#> 4  2013     1     1     544           545          -1      1004          1022
#> 5  2013     1     1     554           600          -6       812           837
#> 6  2013     1     1     554           558          -4       740           728
#> # ... with 3.368e+05 more rows, and 11 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 <dtm>
```