Dplyr Exercises

Dependencies

```
library(nycflights13)
library(dplyr)
library(lubridate)
```

Simple operations

Select

Use the dataset flights.

1. Select all variables whose name contains the word "time"
2. Remove from the previous dataset the variable "arr_time" and "sched_arr_time"

```
data("flights")
```

Filter, Arrange

Use the dataset airports

1. Extract the information of the "Elizabethton Municipal Airport"
2. Extract the information of the airport whose latitude is above 35 and longitude is below -82. How many are them? Order the airports by name
3. Extract the information of the airports in the following time zones: "America/New_York", "America/Chicago", "America/Los_Angeles". Order the airports by latitude and longitude, in decreasing order

```
data("airports")
```

Mutate

Use the dataset `planes`

1. Add a new variable called "age", that shows the age in years of each plane
2. Add a new variable called "if_big", that has value TRUE if the number of seats is above 100, FALSE otherwise

```
data('planes')
```

Multiple operations on single dataset - Analysis pipeline

Answer the following questions:

1. Consider the airports. How many airports there are for each time zone, and what is their average altitude in meters? Sort by number of airports in decreasing order. (1 foot = 0.3048 meters)

```
data("airports")
```

2. Consider the planes. For each model, what is the average number of seats and the average age? Show only the models that have more than 100 planes
3. Consider the flights. What's the average delay at arrival for each route (itinerary with the same origin and destination)? What are the top 3 routes for average arrival delay? Compute also the delay at arrivals 0.25 and the 0.75 quantiles.

Multiple operations on multiple datasets - Analysis pipeline

Check the following hypothesis:

1. The airports that accumulate more arrival delay are located in the Los Angeles time zone.
2. Some weather conditions strongly influence the departure delay. Which ones, if any? Check the weather dataset, and consider only departure delay below 200 minutes.

```
data("weather")
```