R3 Exercises: Joining Tables and Grammer of Graphics

INSERT HERE YOUR NAME

2023-11-30

Table of contents

Packages used in this notebook:

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

---

1 Joining baseball data

Use the data from the `{Lahman}` package. A description of the package is available at CRAN

1.1 Players with 300 home-runs and 300 stolen bases (3P)

Name every player in baseball history who has accumulated at least 300 home runs (HR) and at least 300 stolen bases (SB). (Use the `Batting`, and `People` data from the `Lahman` package).

1.2 Players with 50 home-runs in 3 different years (3P)

Name every player in baseball history who has scored 50 home runs in at least 3 different years. (Use the `Batting`, and `People` data from the `Lahman` package).

1.3 Teams with the highest number of home-runs per player (3P)

Identify the names of the 5 teams in baseball history with the highest number of home runs per player. (Use the `Batting`, and `Teams` data from the `Lahman` package).

2 Joining fligths data

Use the `{nycflights13}` package to answer the following questions. A description of the package is available at CRAN.

2.1 How many different planes? (5P)

Use the `flights` and `planes` tables (in the `{nycflights13}` package) to answer the following questions.

2.1.1 Primary keys

- Does `planes` table have a primary key?
- Can you identify a primary key in `planes` table?
- Show that `tailnum` is not a primary key in `flights` table.
- How many different planes departed from "JFK" in 2013?

2.1.2 Join tables the flight and planes tables.

Use and compare inner/right/left/full join to combine the `flights` data with the `planes` data.

- How many of the planes in the planes table departed from "JFK" in 2013?
- Have planes departed from "JFK" that are not in the planes table?
- How many of the different ***planes in the `planes` table*** flew from New York City in 2013?
- Which join commands yield correct/wrong results and why?

Inner Join

Right Join

Left Join

Full Join

2.2 The oldest plane that departed from New York Airport (3P)

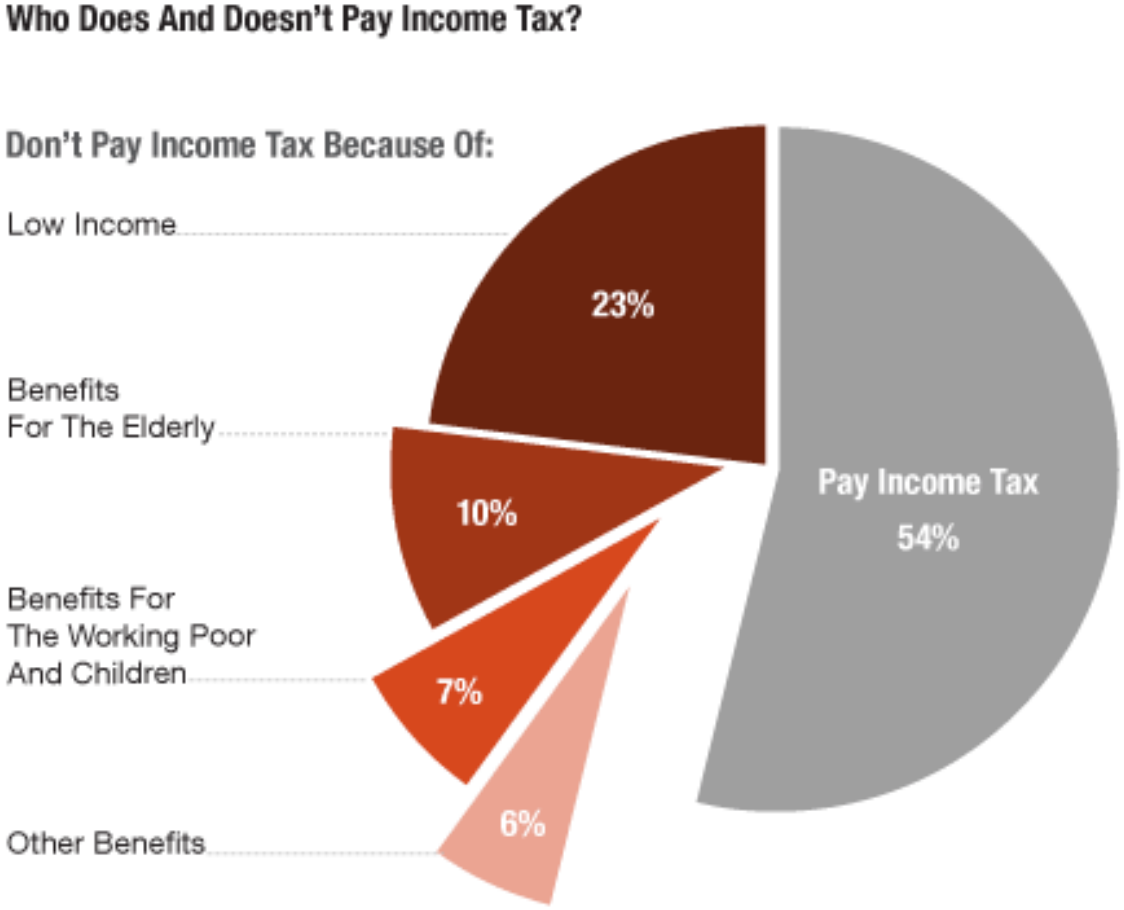
What is the oldest plane (specified by the `tailnum` variable) that flew from New York City airports in 2013? Use the `flights` and `planes` tables (in the `nycflights13` package). Rename `year` in `planes` to `plane_year` to avoid collision with `year` in `flights`.

3 Grammer of Graphics

3.1 Analysis of a graph (4P)

Analyze the following graphical data presentation by considering the following points:

a. Identify the visual cues, coordinate system, and scale(s).
b. How many variables are depicted in the graph?
c. Explicitly link each variable to a visual cue that you listed above.
d. Critique this data graphic using the taxonomy described in lecture slides.



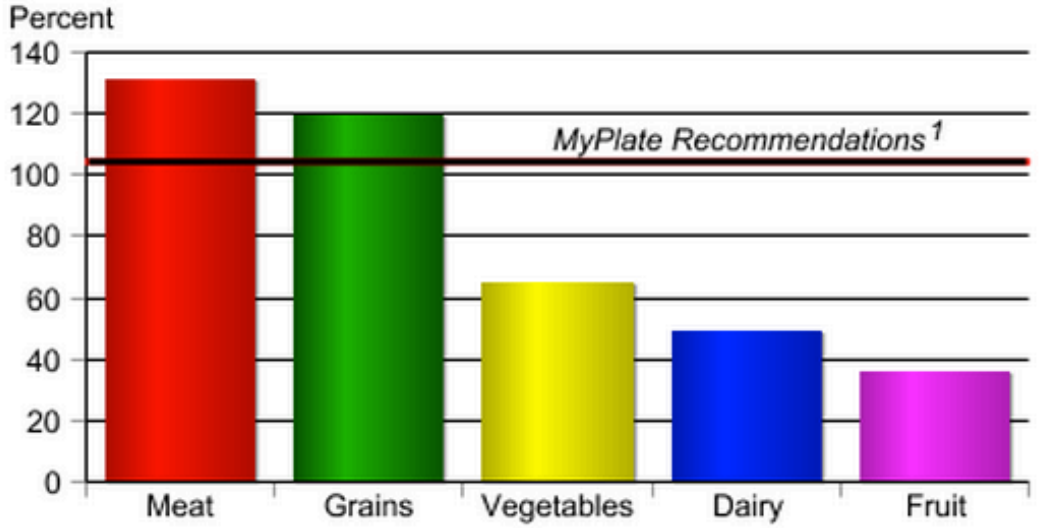
3.2 Improve a graphical data presentation (4P)

Analyze the following graph by considering the following points:

a. Critically assess the designer's choices for the graphical data presentation.
b. Suggest improvements for the graph.

American diets are out of balance with dietary recommendations

In 2012, Americans consumed more than the recommended share of meat and grains in their diets but less than the recommended share of fruit, dairy, and vegetables



Note: Rice and durum flour data were discontinued and thus are not included in the grains group. Food availability data serve as proxies for food consumption.

[1]Data based on a 2,000-calorie diet.

Source: Calculated by ERS/USDA based on data from various sources (see Loss-Adjusted Food Availability Documentation). Data as of February 2014.

Chart data