R5 Exercises: Graphing data variables

Contents

Packages used in this notebook:

---

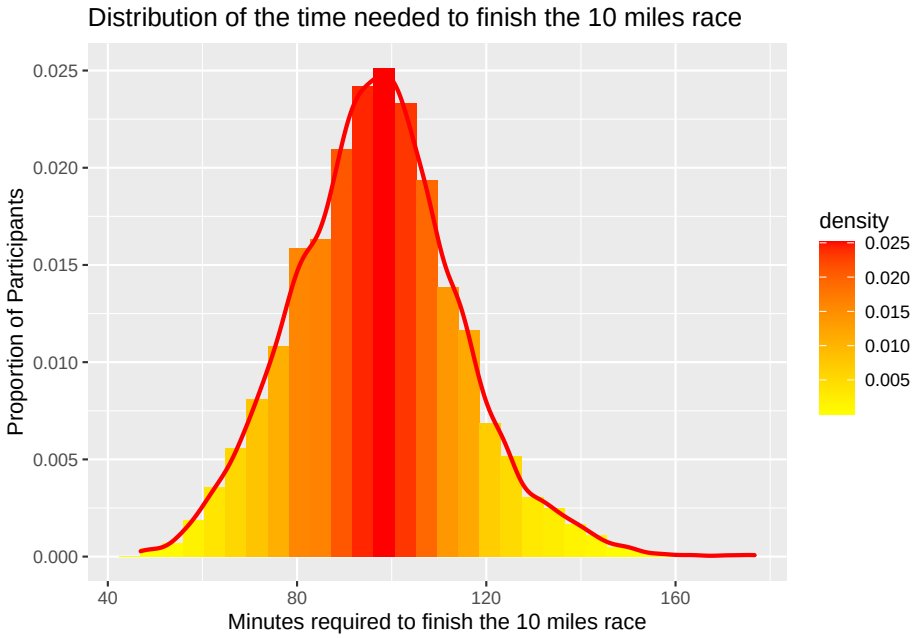
1 Visualization `TenMileRace` data

Use the two variables `time` and `sex` from the `TenMileRace` data in the `mosaicData` package. Choose a suitable visualization method for *each* of them and also for their relationship (create three figures in total). Choose a "Brewer" color palette (using `scale_color_brewer()`, `scale_fill_brewer()`, `scale_color_distiller()` or `scale_fill_distiller()`).

```
## Rows: 8,636
## Columns: 5
## $ state <fct> VA, MD, VA, MD, MD, VA, VA, VA, VA, PA, MD, MD, MD, VA, MD, VA, ~
## $ time  <int> 6060, 4515, 5026, 4229, 5293, 6234, 6286, 6335, 4098, 4855, 5351~
## $ net   <int> 5978, 4457, 4928, 4229, 5076, 5968, 5928, 5939, 4077, 4634, 5105~
## $ age   <int> 12, 13, 13, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 15, 16, 16, ~
## $ sex   <fct> M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M~
```

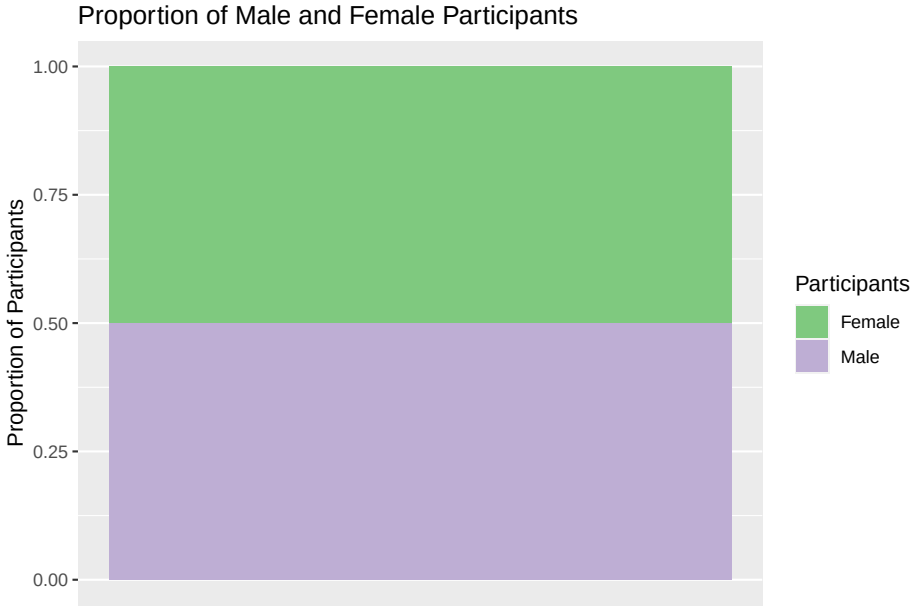
1.1 Visualise variable `time` (3P)

Choose a suitable visualization method for the variable `time` and interpret the diagram.



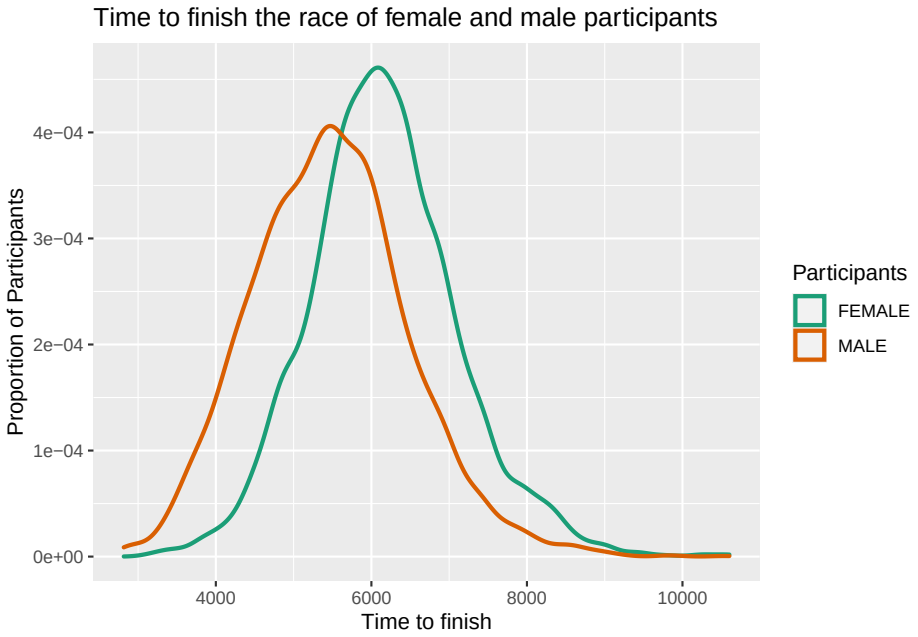
1.2 Visualise variable `sex` (3P)

Choose a suitable visualization method for the variable `sex` and interpret the diagram.



1.3 Visualise the relation between the two variables `time` and `sex` (3P)

Choose a suitable visualization method for the relation between `time` and `sex` and interpret the diagram.



2 Graph flights data

Plot the number of trips per month for the plane with the most flights from New York "JFK" airport. Use the `nycflights13` package. A description of the package is available at https://cran.r-project.org/web/packages/nycflights13/nycflights13.pdf.

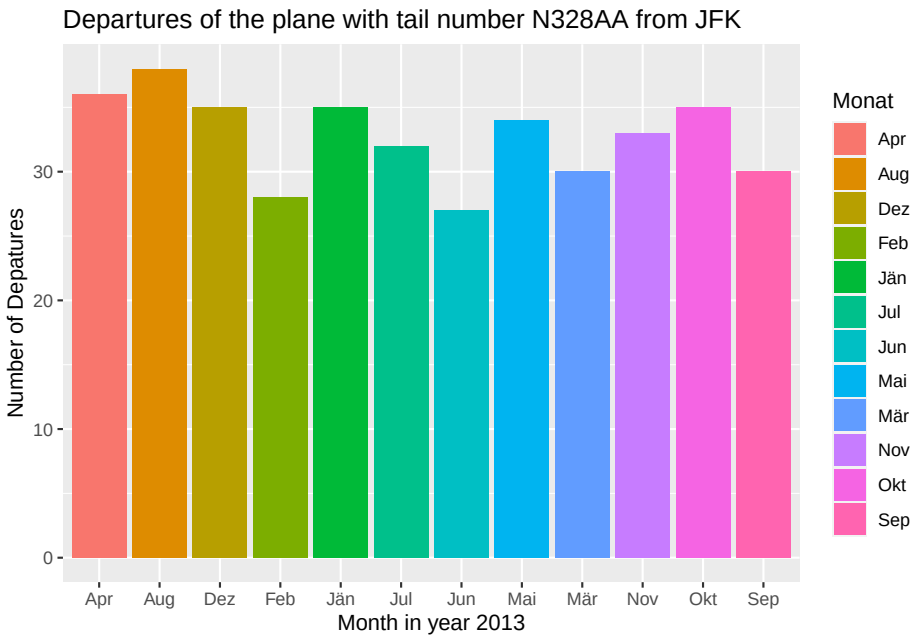
2.1 Identify the `tailnum` of the plane with the most departures (3P)

Identify the plane (specified by `tailnum` in the `flights` data frame) that traveled the most times from New York City ("JFK") airports in 2013 and assign the `tailnum` of this plane to the variable `id_tailnum`.

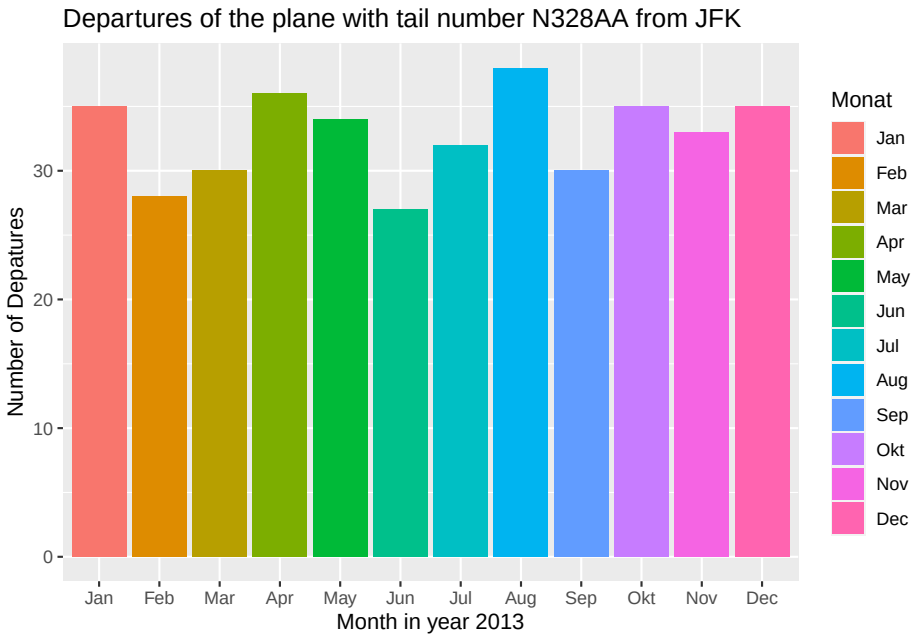
```
## [1] "N328AA"
```

2.2 Graph the number of trips per month (5P).

Plot of the number of trips per month from New York for the plane identified by `id_tailnum`. Include the info in `id_tail_num` into the title of the graph. Use the command `Monat=as.factor(format(time_hour,"%b"))` to extract `Monat` from the variable `time_hour`.



Use the command `fct_reorder(Monat,month)` to reorder the factor `Monat` by the variable `month` (instead of alphabetical ordering of the factor).



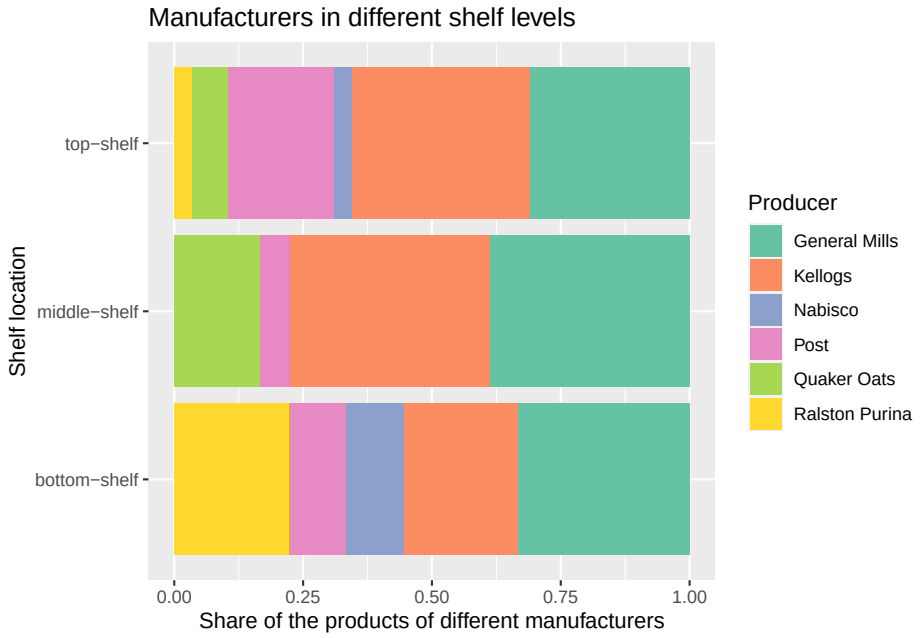
3 Visualise dietary data

Use the code `data("UScereal", package = "MASS")` for the `UScereal` data from the `MASS` package. See https://www.rdocumentation.org/packages/MASS/versions/7.3-53/topics/UScereal for details. Adjust the Manufacturer in `mfr` (represented by its first initial): G=General Mills, K=Kelloggs, N=Nabisco, P=Post, Q=Quaker Oats, R=Ralston Purina and the display shelf in `shelf` (1, 2, or 3, counting from the floor)

into `bottom-shelf`, `middle-shelf` and `top-shelf`. Visualize the relationship of calories, sugars and fat, additionally, highlight whether the product has been enriched with vitamins.

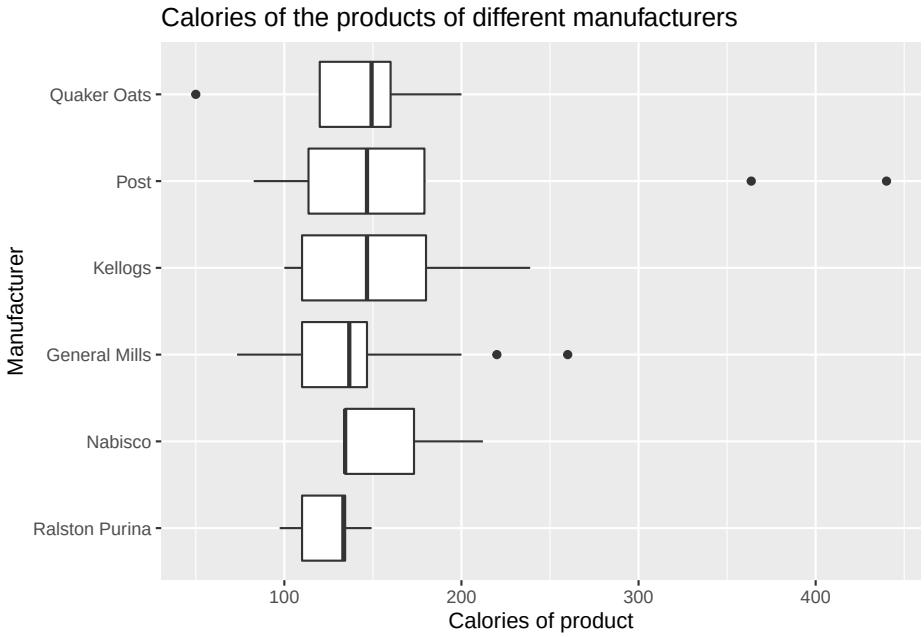
3.1 Distribution of producers (4P)

Visualize the distribution of manufacturers among the shelves. Use appropriate titles and legends.



3.2 Distribution of calories for different producers (4P)

Visualize the amount of calories in the products of different manufacturers with a boxplot and a violin plot. Reorder the manufacturers with the command `fct_reorder(mfr, calories)` in the graph. Use appropriate titles and legends.


Calories of the products of different manufacturers
Quaker Oats
Post
Kellogs
General Mills
Nabisco
Ralston Purina
Manufacturer
100
200
300
400
Calories of product


Calories of the products of different manufacturers
Quaker Oats
Post
Kellogs
General Mills
Nabisco
Ralston Purina
Manufacturer
100
200
300
400
Calories of product