

Demo for Kerstin

Table of contents

Purpose	1
What is the advantage of an R workflow?	1
Figures in ggplot	4

Purpose

A small demonstration of R with our data.

What is the advantage of an R workflow?

We are writing at the moment in a file called a quarto markdown file (.qmd). It is a lot like a dofile in Stata, but it can take normal text in addition to code comments, and when you produce output from your code, it can be shown next to the code that produced it.

The advantage of this workflow is that it is highly reproducible.

For example, say we have a list of people's ages.

```
library(tidyverse) # for plotting and data analysis
library(gt) # for tables
theme_set(theme_light())

ages <- tibble(age = rnorm(n = 100, mean = params$mean_age, sd = 20))

ages
```

```
# A tibble: 100 x 1
  age
<dbl>
1 21.5
2 19.7
3 41.4
4 20.1
5 34.1
6 -2.56
7 13.0
8 18.8
9 38.1
10 34.2
# ... with 90 more rows
```

We can make a quick histogram of the ages using `ggplot` and a summary table using `gt`.

```
ggplot(
  ages, # the object we want to plot
  aes(age) # the variable we want to plot
) +
  geom_histogram(bins = 10) + # the kind of plot we want
  labs(x = "Age",
       y= "Number of observations")
```

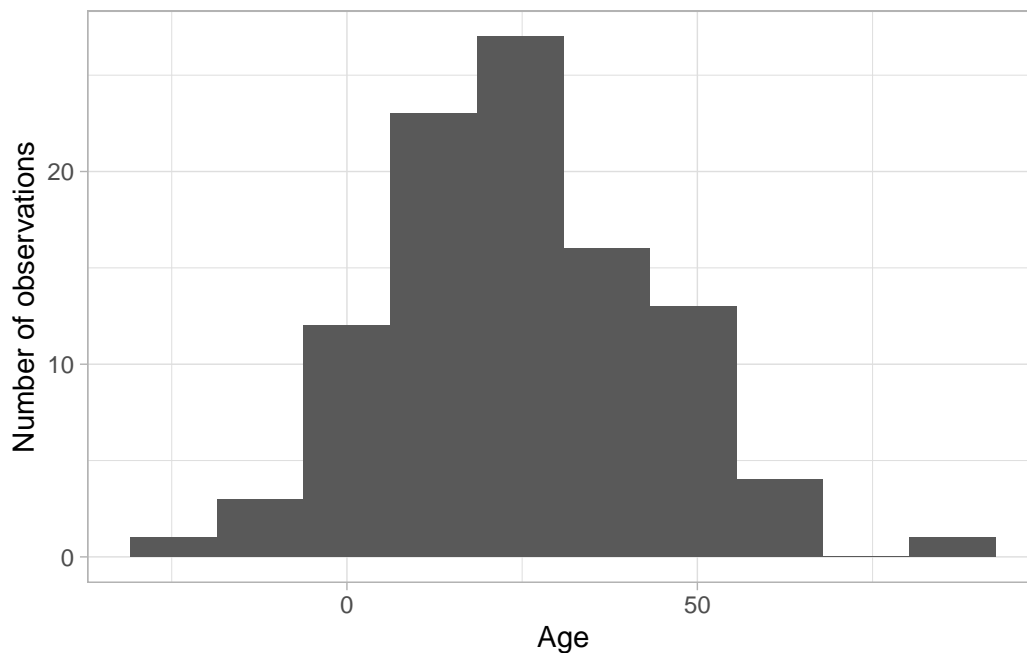


Figure 1: Histogram of ages

```
ages %>%
  summarise(
    `Mean age` = mean(age),
    `Numbe of obs` = n(),
    `Standard dev` = sd(age)
  ) %>%
  pivot_longer(everything(), names_to = "Summary stat", values_to = "Value") %>%
  gt() %>%
  fmt_number(columns = Value, decimals = 2)
```

Table 1: Ages

Summary stat	Value
Mean age	24.68
Numbe of obs	100.00
Standard dev	18.97

If we are told we had the wrong data, and in fact that the mean of the ages is 25 rather than 50, we can reproduce the same report by changing our data (in this case the parameter called

mean_age to 50), and we don't have copy and paste the results into our word document - they just automatically populate this document.

Figures in ggplot