

Data Visualisation with ggplot2

Communicating your data

Daniela Palleschi

2023-04-13

Table of contents

Data communication	2
ggplot2: layered grammar of graphics	2
Plotting reading times	2
Load packages and data	3
Take a look	3
Add labels	4
Add	5
Add condition	6
Distributions	8
Histogram	8
Density plots	8
Grouped density plots	8
Scatterplots	8
Summary statistics	8
Boxplots	8
Violin plots	8
Violin boxplots	8
Interaction plots	8
Facets	8
Resources	9

```

## play sound if error encountered
### from: https://sejohnston.com/2015/02/24/make-r-beep-when-r-markdown-finishes-or-when-i
options(error = function(){ # Beep on error
  beepr::beep(sound = "wilhelm")
  Sys.sleep(2) #
})
## and when knitting is complete
.Last <- function() { # Beep on exiting session
  beepr::beep(sound = "ping")
  Sys.sleep(6) # allow to play for 6 seconds
}

# Create references.json file based on the citations in this script
# make sure you have 'bibliography: references.json' in the YAML
rbbt::bibt_update_bib("_data_viz.qmd")

```

Wrote 2 references to './references/references.json'

```

knitr::opts_chunk$set(eval = T, # change this to 'eval = T' to reproduce the analyses; mak
  echo = T, # 'print code chunk?'
  message = F, # 'print messages (e.g., warnings)?'
  error = F,
  warning = F)

```

Data communication

ggplot2: layered grammar of graphics

- ggplot2 is part of the tidyverse (like dplyr)
 - uses a layered grammar of graphics
 - i.e., we build layers

Plotting reading times

- reading times are (usually) continuous variables

- as are e.g., reaction times
- they are truncated at 0, meaning they cannot have negative values
 - because of this, they tend to have a *skewed distribution*

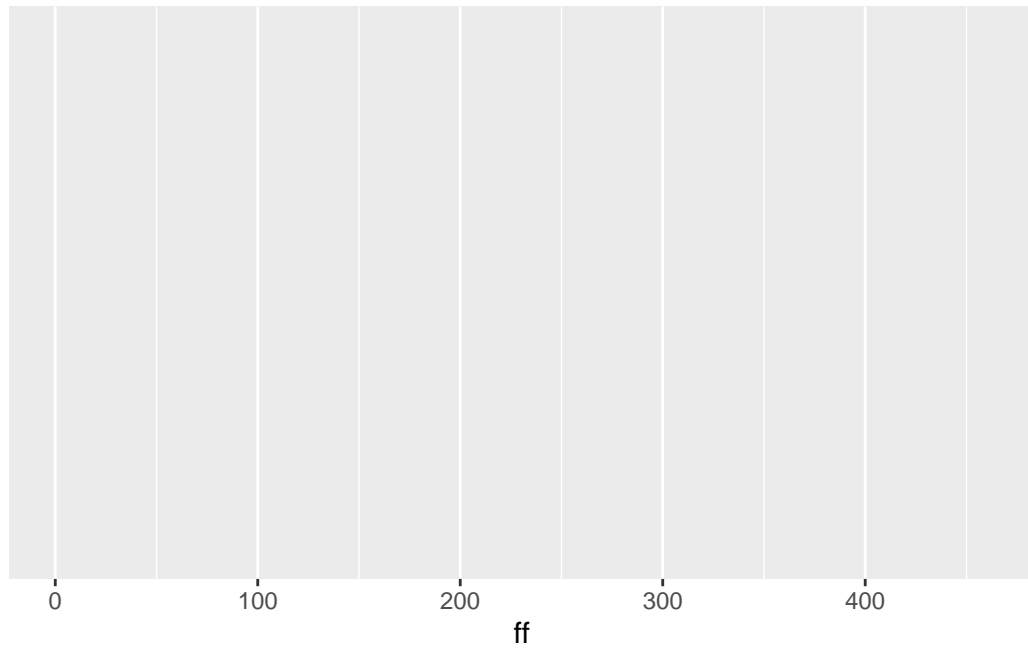
Load packages and data

```
# load tidyverse
library(tidyverse)

# load data
df_lifetime <- readr::read_csv(here::here("data/tidy_data_lifetime_pilot.csv"),
                               # for special characters
                               locale = readr::locale(encoding = "latin1")
                               ) |>
  filter(type == "critical", # only critical trials
         region_n == "2")
```

Take a look

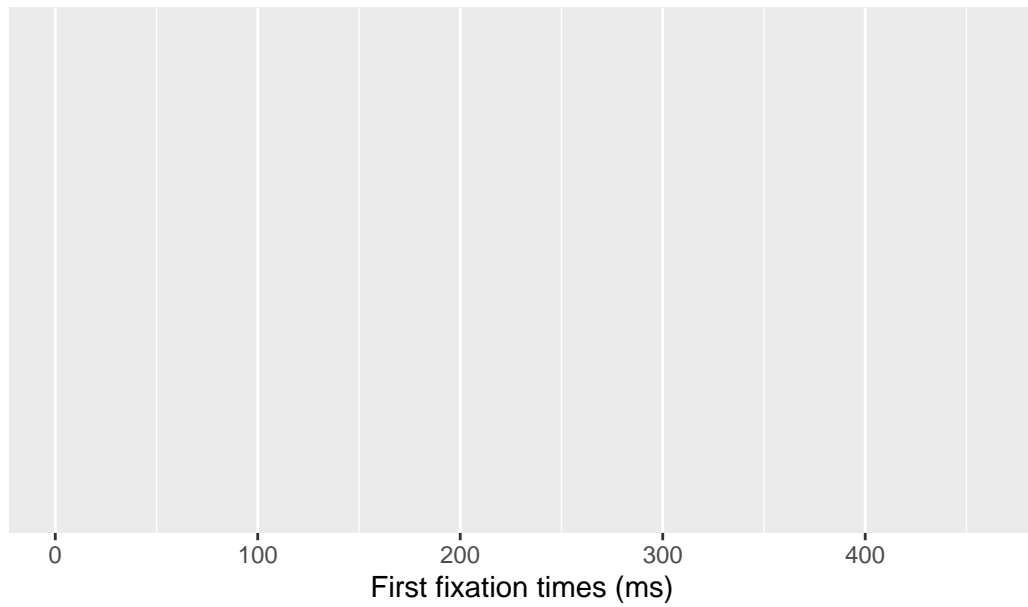
```
df_lifetime |> ggplot(aes(ff))
```



Add labels

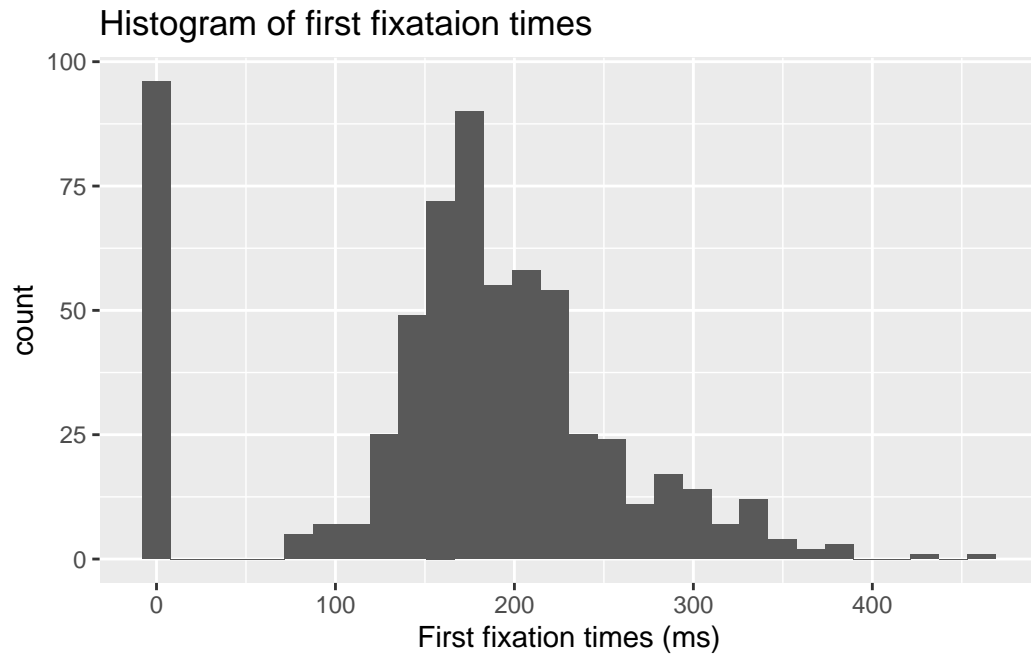
```
df_lifetime |> ggplot(aes(ff)) +  
  labs(title = "Histogram of first fixataion times",  
        x = "First fixation times (ms)")
```

Histogram of first fixataion times



Add

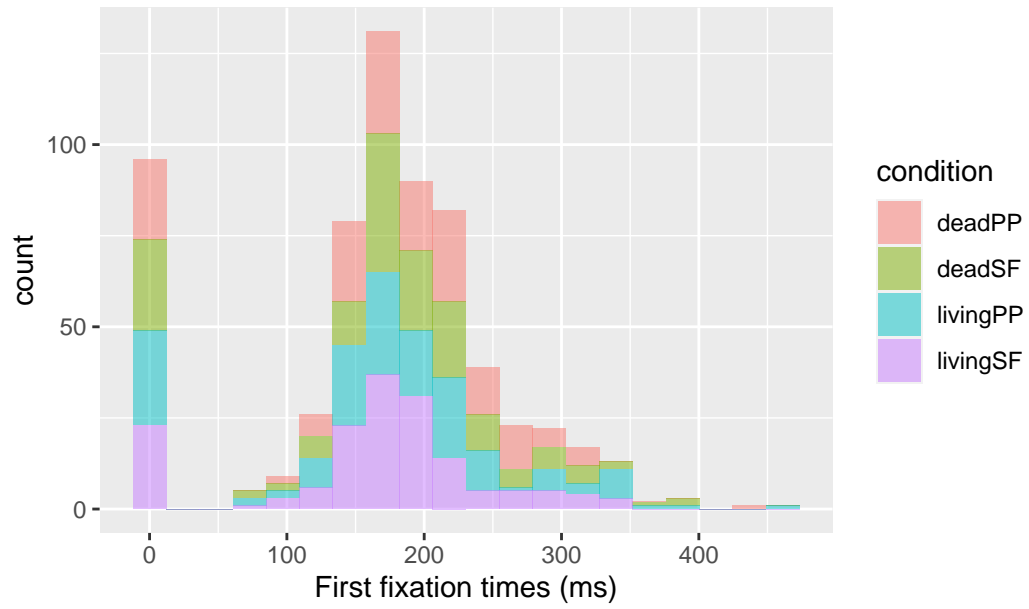
```
df_lifetime |> ggplot(aes(ff)) +  
  labs(title = "Histogram of first fixataion times",  
        x = "First fixation times (ms)") +  
  geom_histogram()
```



Add condition

```
df_lifetime |> ggplot(aes(ff, fill = condition)) +  
  labs(title = "Histogram of first fixataion times",  
        x = "First fixation times (ms)") +  
  geom_histogram(alpha=.5, bins = 20)
```

Histogram of first fixataion times



Distributions

Histogram

Density plots

Grouped density plots

Scatterplots

Summary statistics

Boxplots

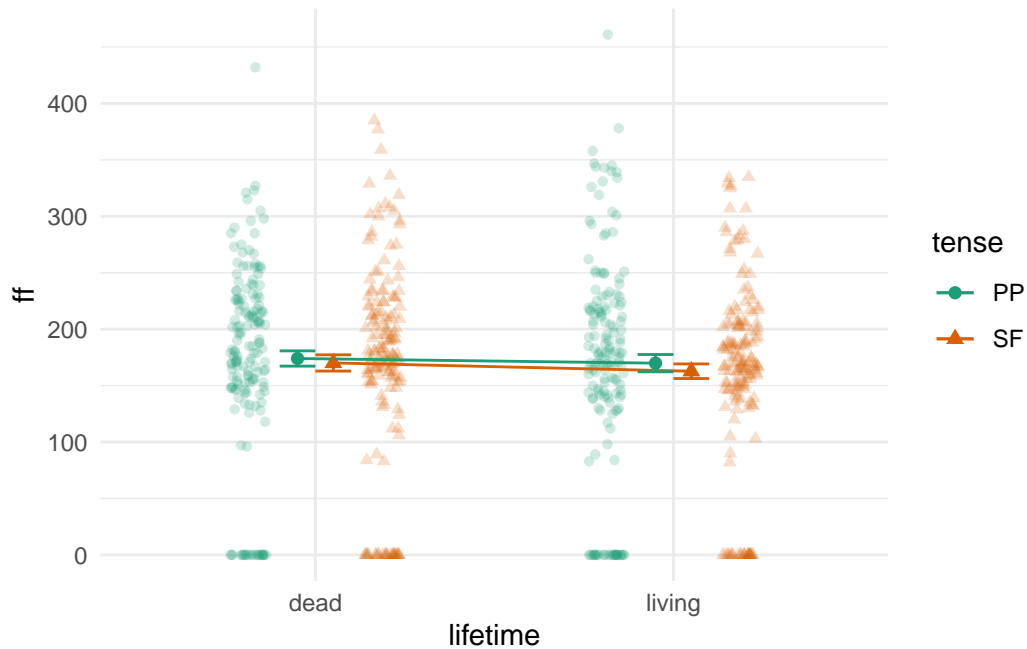
Violin plots

Violin boxplots

Interaction plots

Facets

```
pd = position_dodge(0.2)
ggplot(df_lifetime, aes(x = lifetime, y = ff,
                        group = tense, shape = tense)) +
  # adds raw data points in each condition
  geom_point(aes(colour = tense), position = position_jitterdodge(.1), alpha = .2) +
  # add lines to connect each participant's data points across conditions
  # geom_line(aes(group = item_id, colour = tense), position = position_jitterdodge(.1), a
  # add data points representing cell means
  stat_summary(fun = "mean", geom = "point", size = 2, position = pd, aes(colour = tense))
  # add lines connecting cell means by condition
  stat_summary(fun = "mean", geom = "line", position = pd, aes(colour = tense)) +
  # add errorbars to cell means
  stat_summary(fun.data = "mean_se", geom = "errorbar",
              width = .2, position = pd, aes(colour = tense)) +
  # change colours and theme
  scale_color_brewer(palette = "Dark2") +
  theme_minimal()
```

Resources

Nordmann et al. (2022)

Nordmann and DeBruine (2022)

Nordmann, Emily, and Lisa DeBruine. 2022. “Applied Data Skills.” Zenodo. <https://doi.org/10.5281/zenodo.6365078>.

Nordmann, Emily, Phil McAleer, Wilhelmiina Toivo, Helena Paterson, and Lisa M. DeBruine. 2022. “Data Visualization Using R for Researchers Who Do Not Use R.” *Advances in Methods and Practices in Psychological Science* 5 (2): 251524592210746. <https://doi.org/10.1177/25152459221074654>.