How to plot a 'percentage plot' with ggplot2

NOVEMBER 03, 2016

Reading time ~1 minute

At times it is convenient to draw a frequency bar plot; at times we prefer not the bare frequencies but the proportions or the percentages per category. There are lots of ways doing so; let's look at some `ggplot2` ways.

First, let's load some data.

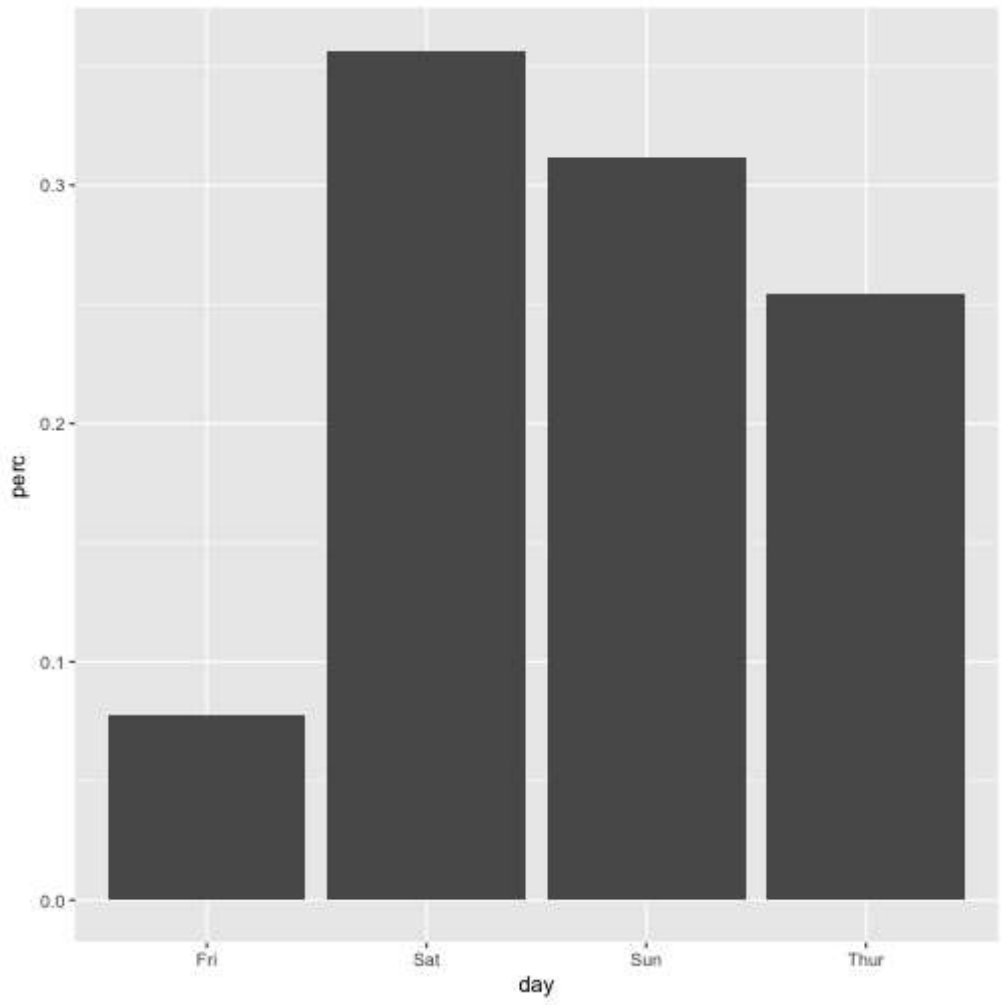
```
data(tips, package = "reshape2")
```

And the typical libraries.

```
library(dplyr)
library(ggplot2)
library(tidyr)
library(scales)  # for percentage scales
```

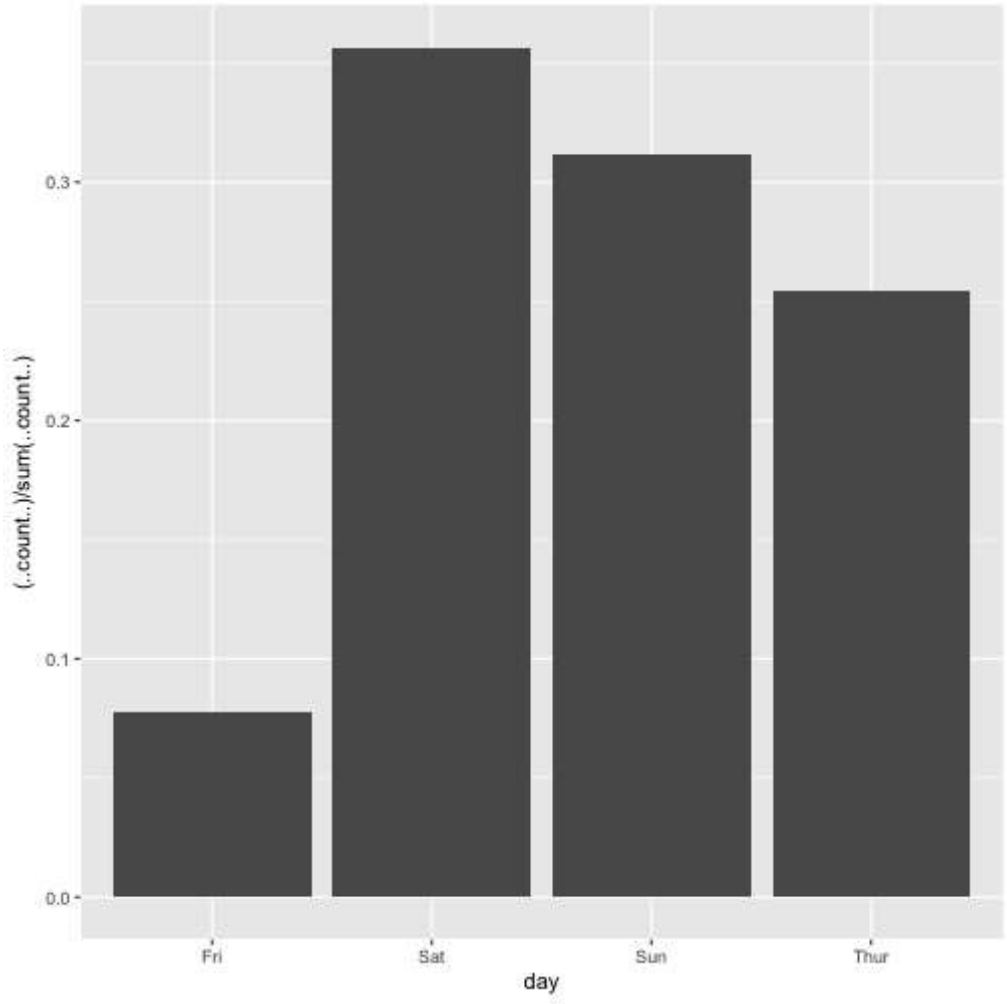
Way 1

```
tips %>%
  count(day) %>%
  mutate(perc = n / nrow(tips)) -> tips2

ggplot(tips2, aes(x = day, y = perc)) + geom_bar(stat = "identity")
```



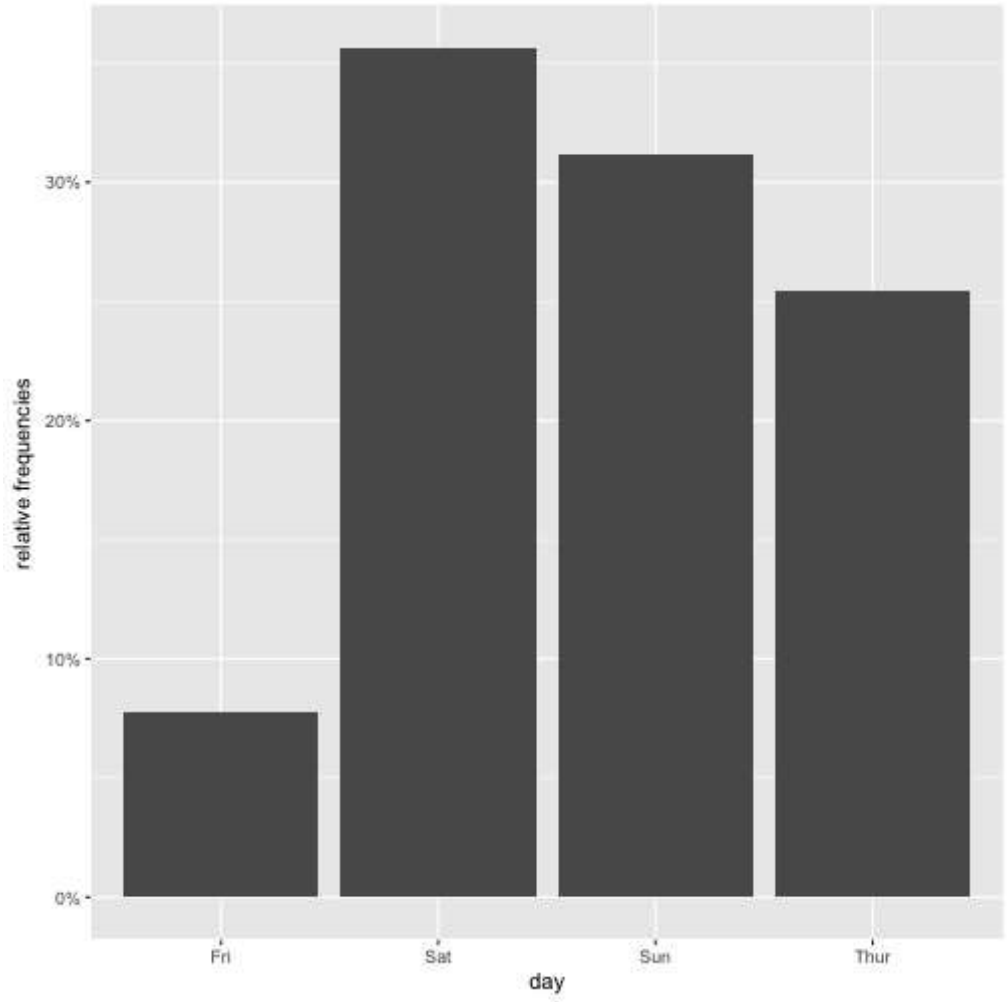
Way 2

```
ggplot(tips, aes(x = day)) +
  geom_bar(aes(y = (..count..)/sum(..count..)))
```



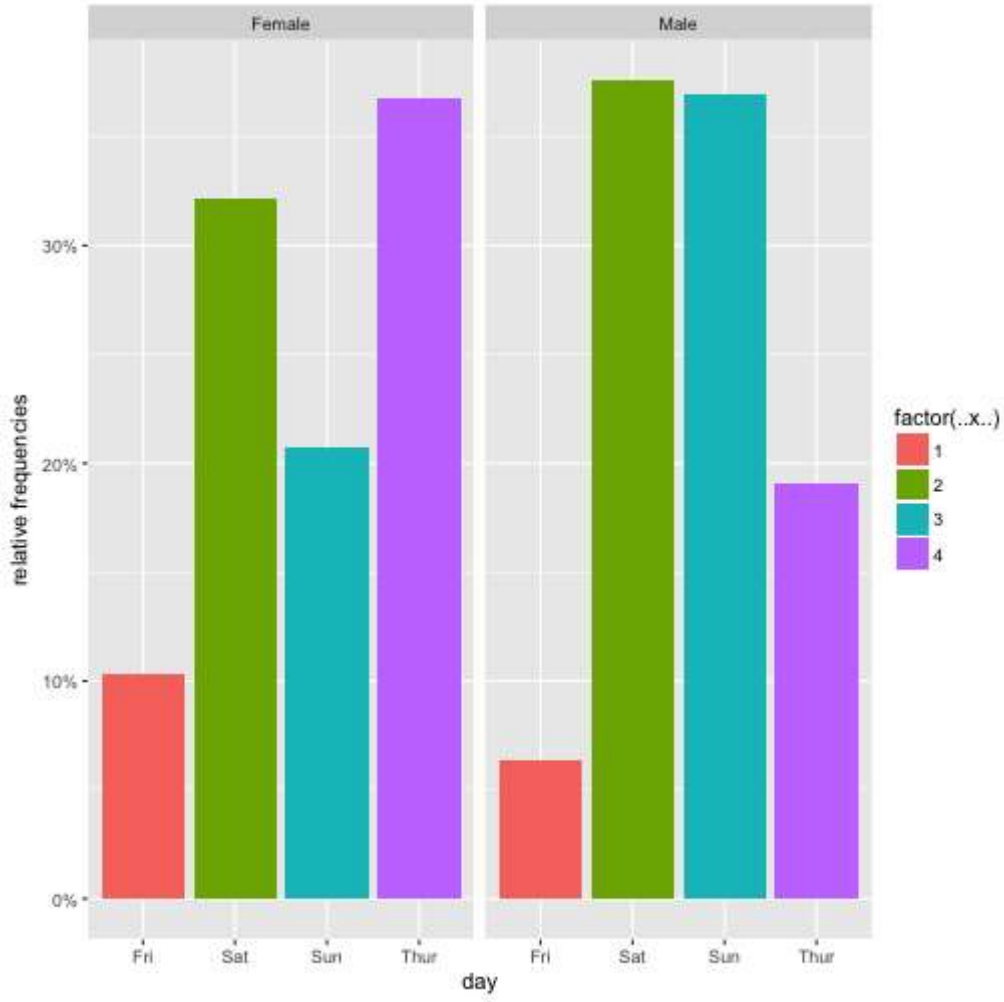
Way 3

```
myplot <- ggplot(tips, aes(day)) +
          geom_bar(aes(y = (..count..)/sum(..count..))) +
          scale_y_continuous(labels=scales::percent) +
  ylab("relative frequencies")

myplot
```



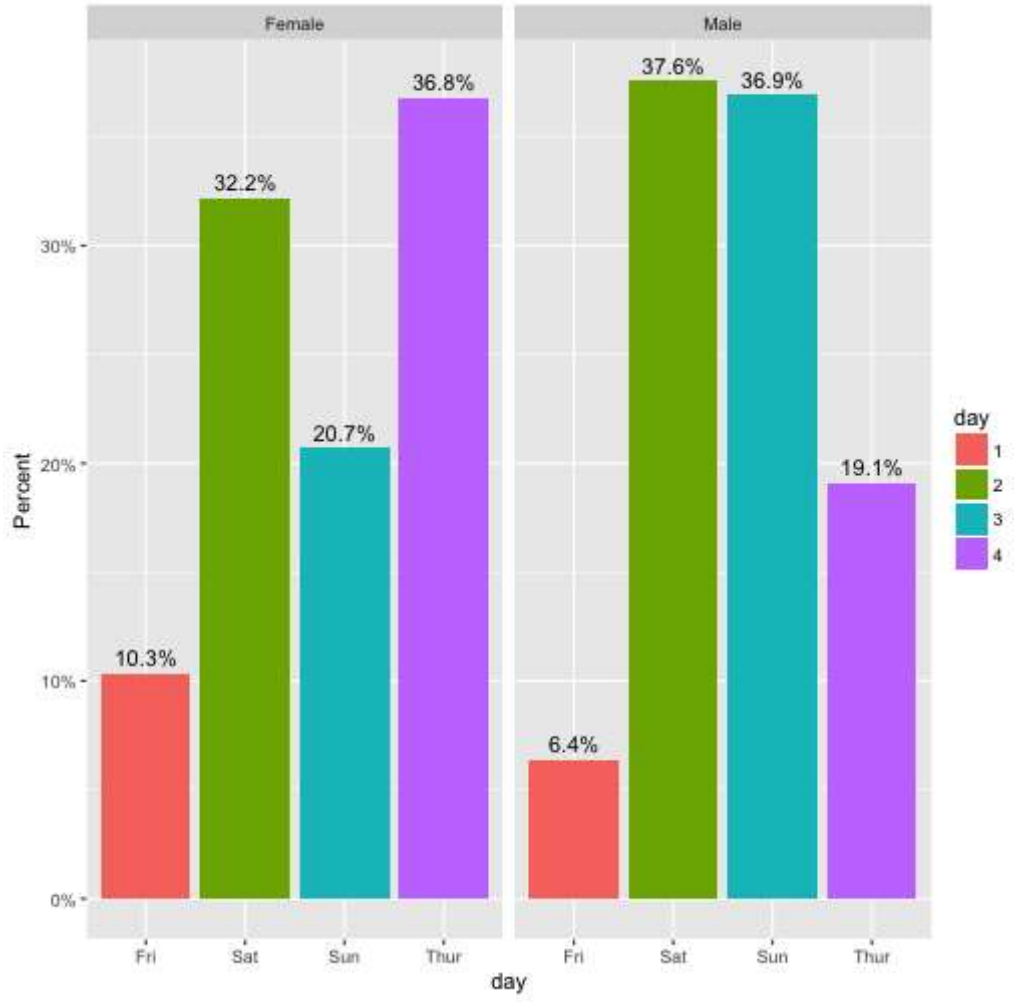
Way 4

```
myplot <- ggplot(tips, aes(day, group = sex)) +
          geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +
          scale_y_continuous(labels=scales::percent) +
          ylab("relative frequencies") +
          facet_grid(~sex)

myplot
```



Way 5

```
ggplot(tips, aes(x= day,  group=sex)) +
    geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +
    geom_text(aes( label = scales::percent(..prop..),
                   y= ..prop.. ), stat= "count", vjust = -.5) +
    labs(y = "Percent", fill="day") +
    facet_grid(~sex) +
    scale_y_continuous(labels = scales::percent)
```



PLOTTING RSTATS

LIKE TWEET +1

24 Comments Sebastian Sauer Blog Login

Recommend 6 Tweet Share Sort by Best

Join the discussion…

LOG IN WITH

OR SIGN UP WITH DISQUS

Name

Merce • a year ago

Hi Sebastian, thanks for this useful and interesting post.

I have been trying all the options, and been able to produce very nice graphs. Just getting in trouble with decimals in the last graph, with awful labels (70.428%, 15.953% and so on).

How would you round the percents on the labels of the last graph? I have been trying with the arguments round, accuracy, but nothing seems to work :(

Thanks a lot in advance!

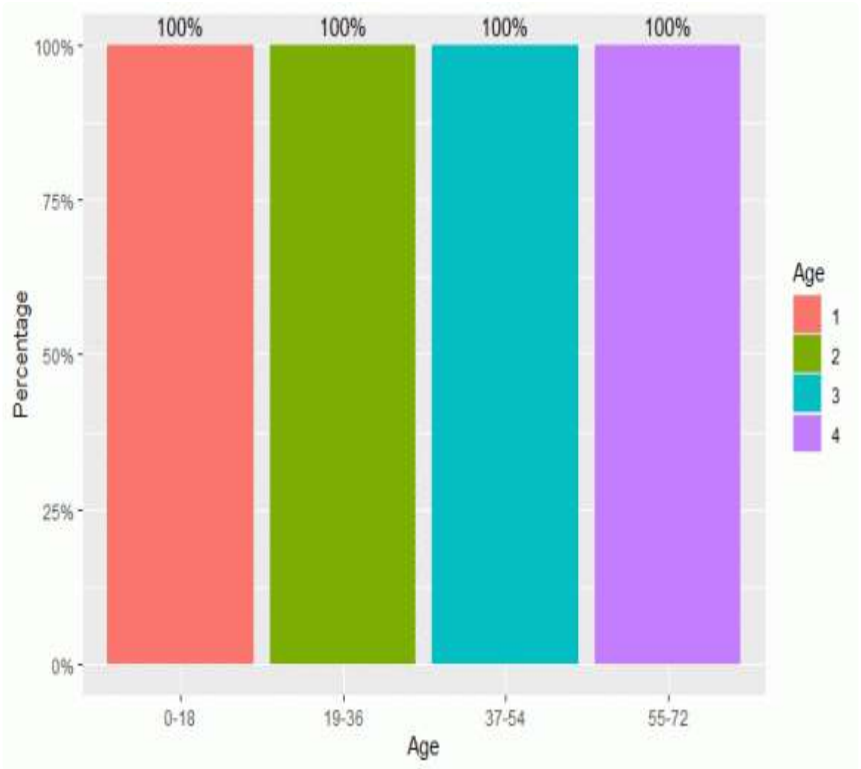
1 ^ | v • Reply • Share ›

elarry ↗ Merce • 5 months ago

`scales::percent(..prop.., accuracy = 0.1)` worked for me - rounded to 1 decimal place.

1 ^ | v • Reply • Share ›

Ankit Singh • 2 months ago



Age	Percentage
0-18	12.22%
19-36	51.11%
37-54	20%
55-72	16.66%

```
ap1 = ggplot(Data1a, aes(x = Age, y = Percentage))+
geom_bar(aes(y = ..prop.., fill= factor(..x..)), stat = "count")+
geom_text(aes(label = scales::percent(..prop..),
y = ..prop..), stat = "count", vjust = -.5)+
labs(y = "Percentage", fill = "Age") +
scale_y_continuous(labels = scales::percent)
```

ap1

I am using this above codes but the result are not correct

Reply • Share ›

Fernando • 4 months ago

Hi Sebastian!
For the way 5, how would you use a named palette for the day colours?
I´m trying

```
ggplot(aes(x = day, group = sex)) +
geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat = "count") +
scale_fill_manual(values = days_colors)
```

but the bars come empty.
Thanks!
Nando

Reply • Share ›

Fernando ➚ Fernando • 4 months ago

Sorry, forgot to map the fill, but still doesn't work

```
ggplot(aes(x = day, group = sex, fill=day)) +
geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat =
"count") +
scale_fill_manual(values = days_colors)
```

Reply • Share ›

Synat • 2 years ago • edited

Dear all,

I am very new to R. I have data which are already in percentage at each timepoint. I could not make a bar graph as percentage as R requires the count. Is there any direct way I can do to make Bar graph directly from my data without going through any transformation process.

Time Cell proliferation

1 hr 34 %

3 hrs 39 %

6 hrs 85%

24 hrs 89%

48 hrs 78%

72 hrs 67%

I wanted to make a bar graph from this data.

thanks,

Reply • Share ›

Nafia Chowdhury • 2 years ago

Hi, I'm trying to plot your last graph such that if I used the above chart, then the friday(female)+friday(male)=100%, and saturday(female)+saturday(male)=100%, etc. How do you suggest I approach that?

Reply • Share ›

MikeSp • 3 years ago

Hello, I am trying "way 3" plot on my data.
myplot <- ggplot(dataframe, aes(marker, group = factor)) +
geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +
scale_y_continuous(labels=scales::percent) +
ylab("relative frequencies") +
facet_grid(~factor)
However, my marker is only pos/neg, so I would like to plot only positive %.
Any advice?

Reply • Share ›

Sebastian Sauer Mod → MikeSp • 3 years ago

Hi Mike, could you provide a reproducible example?

Reply • Share ›

jtr13 • 3 years ago

Another strange behavior, imho, is that Way 4 still works properly (gets the relative frequencies right) if you change + facet_grid(~sex) to + facet_grid(time~sex) without changing the grouping variable. Along these lines, changing group = sex to group = time produces the same plot. Somehow it "knows" to group by both and then calculate the relative frequencies.

Reply • Share ›

Sebastian Sauer Mod → jtr13 • 3 years ago

I think the variable "group" is not really used for computing the frequencies, apparently this is done by whatever "groups" is indicated by facet_grid()

Reply • Share ›

jtr13 • 3 years ago

Thanks, didn't know about ..prop.. I assume fill=factor(..x..) is a workaround since fill = day doesn't work with group = sex. Not sure though why that should be the case; it's not intuitive that those two items would create a conflict.

Reply • Share ›

Sebastian Sauer Mod → jtr13 • 3 years ago

This is an easy way for filling the bars with a user-defined color that works:

```
ggplot(tips, aes(x = day, fill = day)) + geom_bar() +
scale_fill_manual(values = c("red", "blue", "green",
"yellow")) +
facet_grid(~sex)
```

Of course, we can ggplot let ggplot choose the colors:

```
ggplot(tips, aes(x = day, fill = day)) + geom_bar() +
facet_grid(~sex)
```

^ | v • Reply • Share ›

Sebastian Sauer Mod → jtr13 • 3 years ago

Agree. It's kind of strange, not very elegant imho.

^ | v • Reply • Share ›

Gonzalo Franetović Guzmán • 4 years ago

Hi,
Do you know how can I get the Way 5 without decimals?
Thanks!

^ | v • Reply • Share ›

Sebastian Sauer Mod → Gonzalo Franetović Guzmán • 4 years ago

Hi! This worked for me; replace line 3 by this code:
geom_text(aes(label = scales::percent(round(..prop.., 2)),

^ | v 1 • Reply • Share ›

Gonzalo Franetović Guzmán → Sebastian Sauer • 4 years ago

Excelent! Thank you very much Sebastian.

^ | v • Reply • Share ›

GabbyBD • 4 years ago

Hi, where in the ggplot2 documentation can we find more info on ..(function).. ? I've never seen this.

^ | v • Reply • Share ›

Sebastian Sauer Mod → GabbyBD • 4 years ago

Hi, they are not so much documented; for many use cases they are not needed, that's maybe why Hadley chose not to really document them.

But:
For the count and frequency variables, you can refer to Hadley book , page 69:

Both the histogram and frequency polygon geom use stat_bin. This statistic produces two output variables count and density. The count is the default as it is most interpretable. The density is basically the count divided by the total count, and is useful when you want to compare the shape of the distributions, not the overall size. You will often prefer this when comparing the distribution of subsets that have different sizes.

(quoted from: https://stackoverflow.com/q...

Reply • Share ›

GabbyBD ➚ Sebastian Sauer • 4 years ago

Thanks!

Reply • Share ›

Bart • 4 years ago

The bars are always in alphabetical order, so your days aren't in chronological order. Is there a way to customize the order of the x axis? Your example data frame is in alphabetical order to begin with, but I have one that is not. The plot function forces it

Read More

This blog has moved

This blog has moved to Adios, Jekyll. Hello, Blogdown!... Continue reading

Wie gut schätzt eine Stichprobe die Grundgesamtheit?

Published on November 17, 2017

Some thoughts on tidyveal and environments in R

Published on November 16, 2017