ggplot2

McMaster University

Caitlin Simopoulos

2017-06-01

ggplot2

"ggplot2 is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics." -Hadley Whickam

The ‘Grammer of Graphics’

ggplot2 was inspired by the grammer of graphics (hence the ‘gg’ in ggplot2) by Leeland Wilkinson.
Wilkonson *et al.* describe a unique formula for creating and displaying information.

Note: I have not read this book...

ggplot2 relies on the other ‘tidyverse’ packages

The tidyverse can be found at <tidyverse.org>.
Some other packages from the tidyverse you might find useful are:
- dplyr
- tidyr

But we are not here to learn the tidyverse or debate about how much we love and/or hate it...

Free online tutorial

Datacamp offers its introduction to ggplot free!

Click here

How do we reshape data?

```r
library(reshape2)
data("iris")
iris$sample.num <- rownames(iris)
reshapedIris <- melt(iris, id.vars = c("Species", "sample.num"))
```

OK, how would we reshape this data?

```r
data("USArrests")
# first, get information on the dataset...how?
head(USArrests)
```

```
##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58 21.2
## Alaska       10.0     263       48 44.5
## Arizona       8.1     294       80 31.0
## Arkansas      8.8     190       50 19.5
## California    9.0     276       91 40.6
## Colorado      7.9     204       78 38.7
```

Other useful “reshape2” commands...

You should look into:

1. `merge()`
2. `cast()`

Before we start

This is is NO way a complete overview of ggplot2.

This package can do WAY more, but hopefully it gives you an idea on how to start things off.

There is a multitude of tutorials online, so there will always be help if needed :)

OK, let's plot this data!

So that we're all on the same page...

```
USArrests$state <- rownames(USArrests)
reshapedUS <- melt(USArrests, id.vars="state")
head(reshapedUS)
```

```
##        state variable value
## 1    Alabama   Murder  13.2
## 2     Alaska   Murder  10.0
## 3    Arizona   Murder   8.1
## 4   Arkansas   Murder   8.8
## 5 California   Murder   9.0
## 6   Colorado   Murder   7.9
```

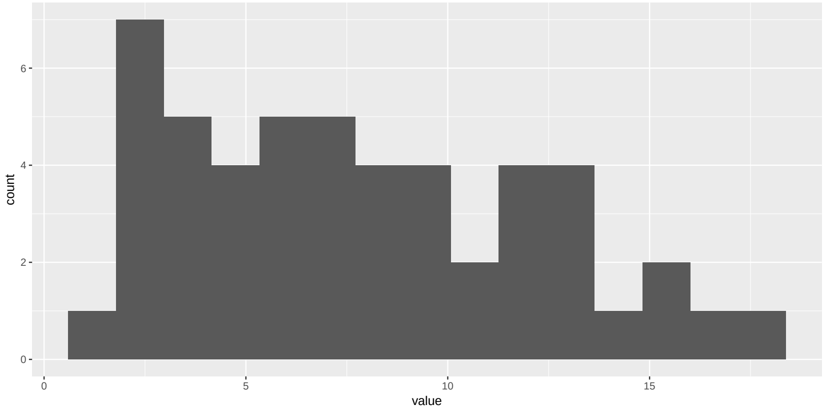
Basic syntax of ggplot2

```r
ggplot(data = yourdata, aes(x=?, y=?, fill = ?)) +
  # tells R that you want to make a plot using ggplot
  geom_line() + # what kind of plot you want
  theme() # where you change font size, colour, etc
```

Histogram

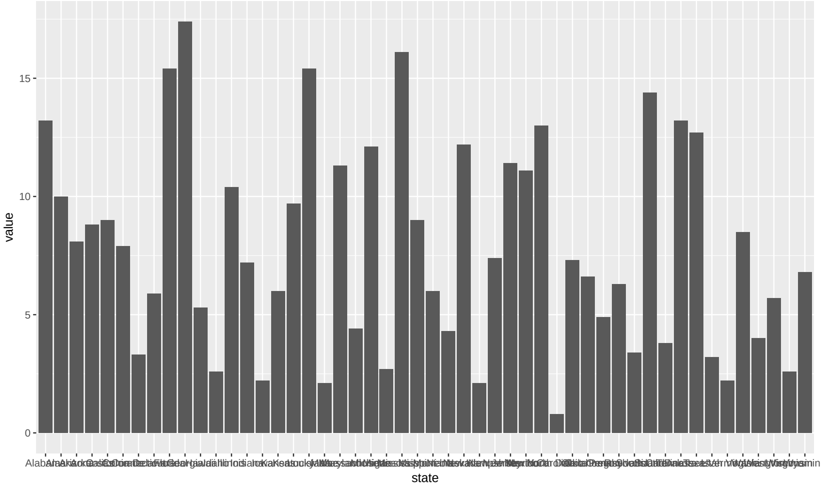
```
library(ggplot2)
murderArrest <- subset(reshapedUS, variable == "Murder")

ggplot(murderArrest, aes(x=value)) +
  geom_histogram(bins = 15)
```



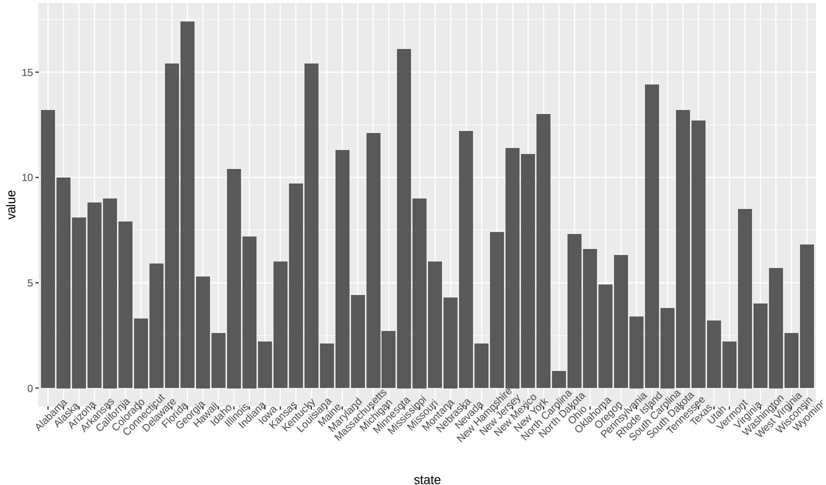
Barchart

```
ggplot () +
  geom_bar(data=murderArrest, aes(x=state, y= value), stat="identity")
```



Barchart

```
ggplot () +
  geom_bar(data=subset(reshapedUS, variable == "Murder"),
           aes(x=state, y= value), stat="identity") +
  theme(axis.text.x = element_text(angle = 45))
```



So many options...

What if we wanted to:

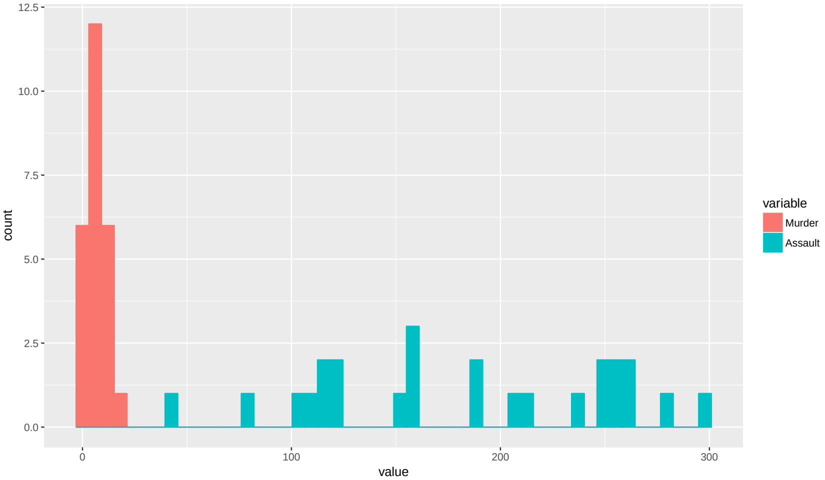
- add an x axis label? y axis label?
- plot “Murder”” and “Assult”
- change the colours?
- add % urban population in there too?
- rotate the bar graph?
- make a box plot?

WE CAN DO ALL OF THIS!!

...but should we?

Murder and assault distribution

```
ggplot(subset(reshapedUS, variable == c("Murder", "Assault")), aes(x=va
  geom_histogram(bins=50, position="identity")
```



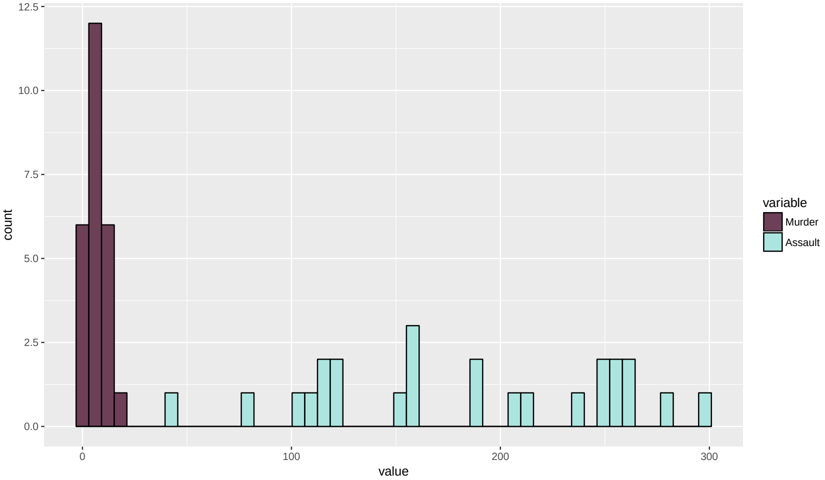
```
# try removing fill = variable
```

Changing colours

- scale_fill_manual() : to use custom colors
- scale_fill_brewer() : to use color palettes from `RColorBrewer` package
- scale_fill_grey() : to use grey color palettes

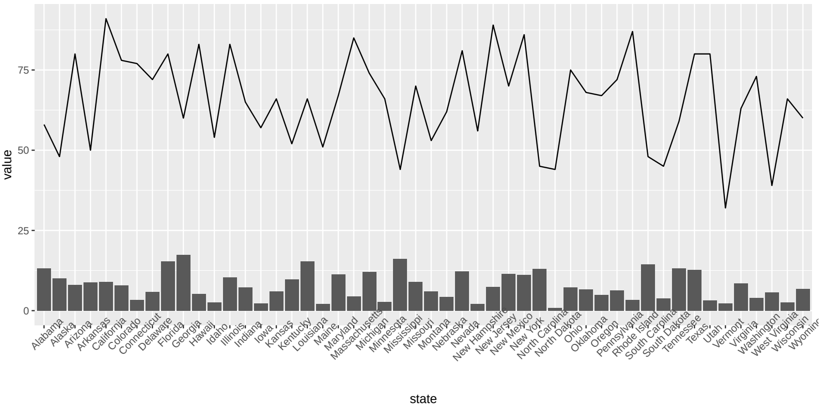
Changing colour example

```
ggplot(subset(reshapedUS, variable == c("Murder", "Assault")), aes(x=va
  geom_histogram(bins=50, position="identity", colour = "black") +
 # scale_color_manual(values=c("#6d4057", "#ace5df")) +
  scale_fill_manual(values=c("#6d4057", "#ace5df"))
```



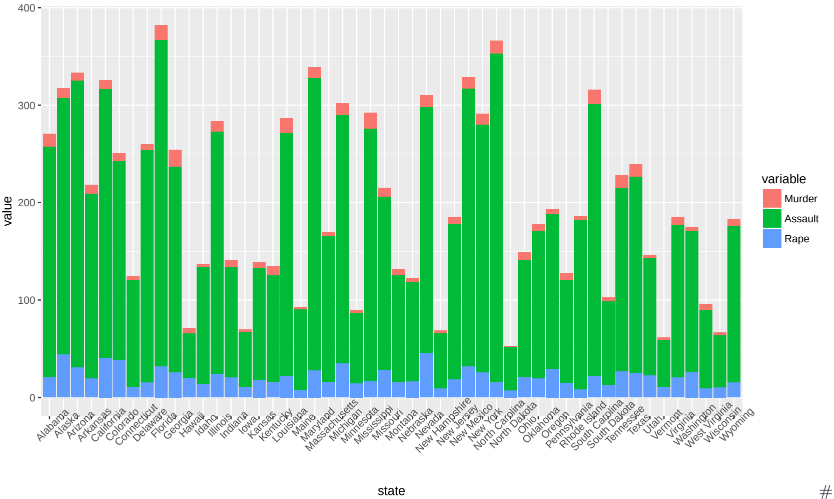
Adding line graph to bar chart

```
ggplot () +
  geom_bar(data=subset(reshapedUS, variable == "Murder"),
           aes(x= state, y= value), stat="identity") +
  geom_line(data = subset(reshapedUS, variable == "UrbanPop"), aes(x=st
  theme(axis.text.x = element_text(angle = 45))
```



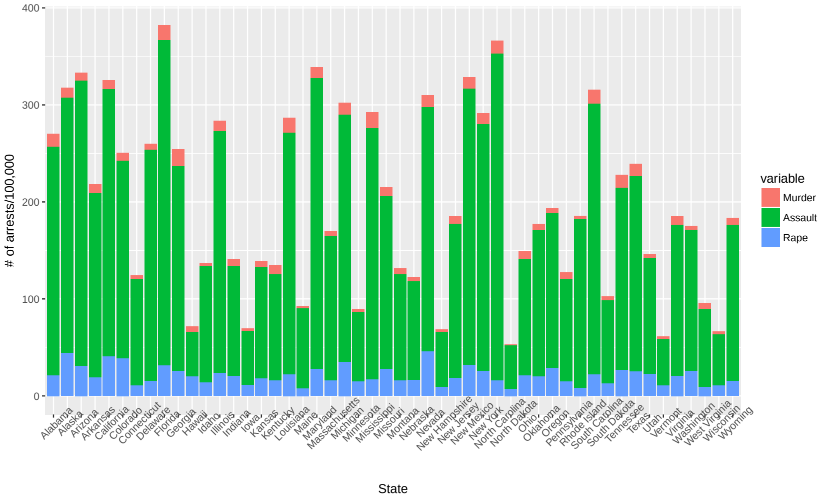
Stacked bar chart with all arrest types

```
barchart <- ggplot () +
  geom_bar(data=subset(reshapedUS, variable != "UrbanPop"),
           aes(x= state, y= value, fill = variable),
           stat="identity") +
  theme(axis.text.x = element_text(angle = 45))
print(barchart)
```



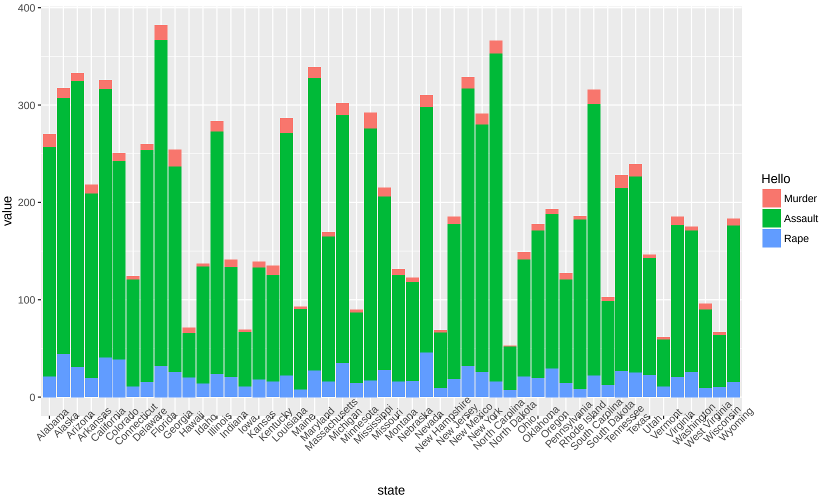
Change labels

```
barchart +
  labs(x = "State", y = "# of arrests/100,000")
```



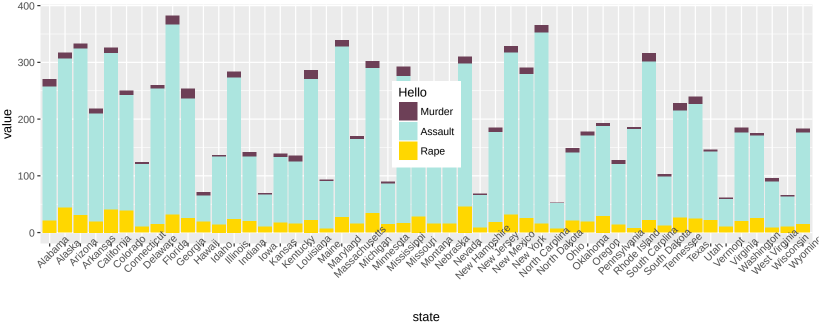
Change legend title

```
barchart +
  guides(fill=guide_legend(title="Hello"))
```



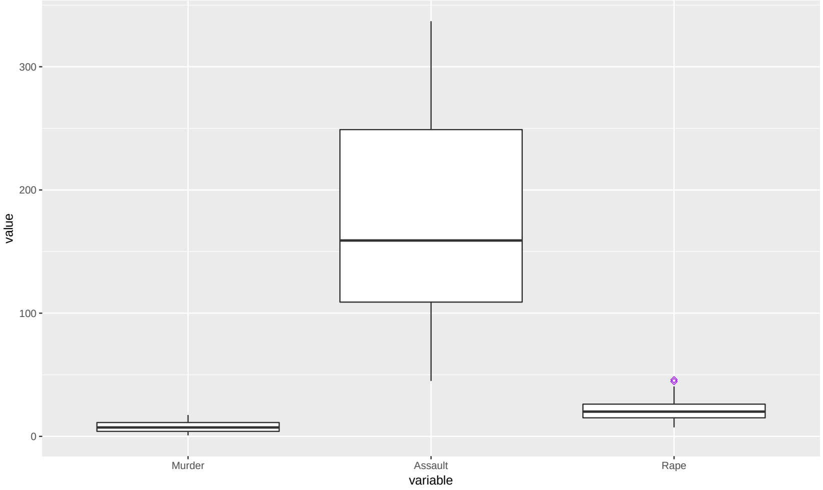
Change legend position and colours

```r
# Position legend in graph,
# where x,y is 0,0 (bottom left)
# to 1,1 (top right)
barchart +
  guides(fill=guide_legend(title="Hello")) +
  theme(legend.position=c(.5, .5)) +
  scale_color_manual(values=c("#6d4057", "#ace5df", "gold")) +
  scale_fill_manual(values=c("#6d4057", "#ace5df", "gold"))
```



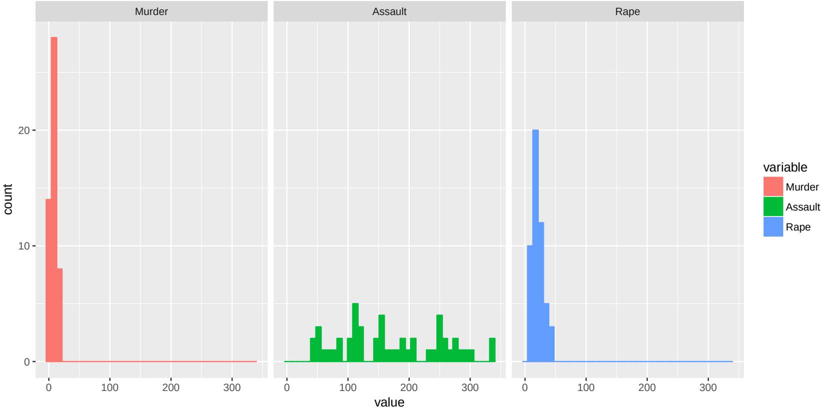
Box plots

```
ggplot(data=subset(reshapedUS, variable != "UrbanPop"),
       aes(x=variable, y=value)) +
   geom_boxplot(outlier.colour = "purple", outlier.shape=5)
```



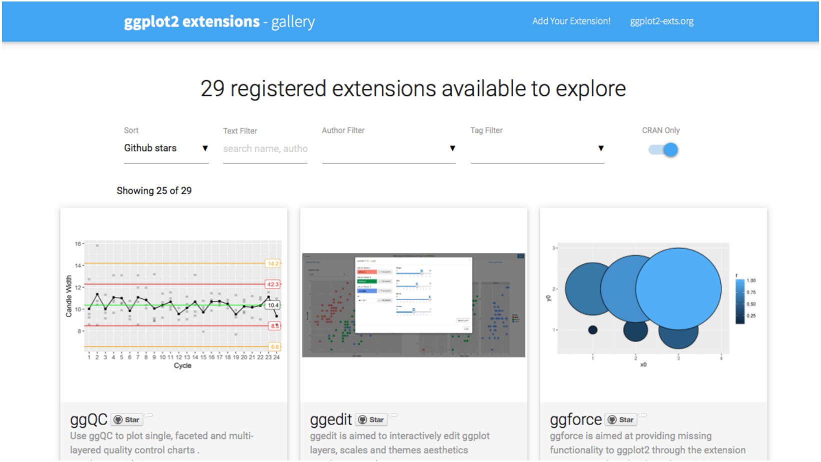
Facets

```
ggplot(data=subset(reshapedUS, variable != "UrbanPop")) +
   geom_histogram(aes(x=value, colour = variable,
                      fill=variable), bins=40) +
  facet_grid(~ variable)
```



ggplot2 extensions

website



ggplot2 extensions - gallery

Add Your Extension! ggplot2-exts.org

29 registered extensions available to explore

Sort: Github stars ▼

Text Filter: search name, autho

Author Filter ▼

Tag Filter ▼

CRAN Only ●

Showing 25 of 29

ggQC ★ Star

Use ggQC to plot single, faceted and multi-layered quality control charts .

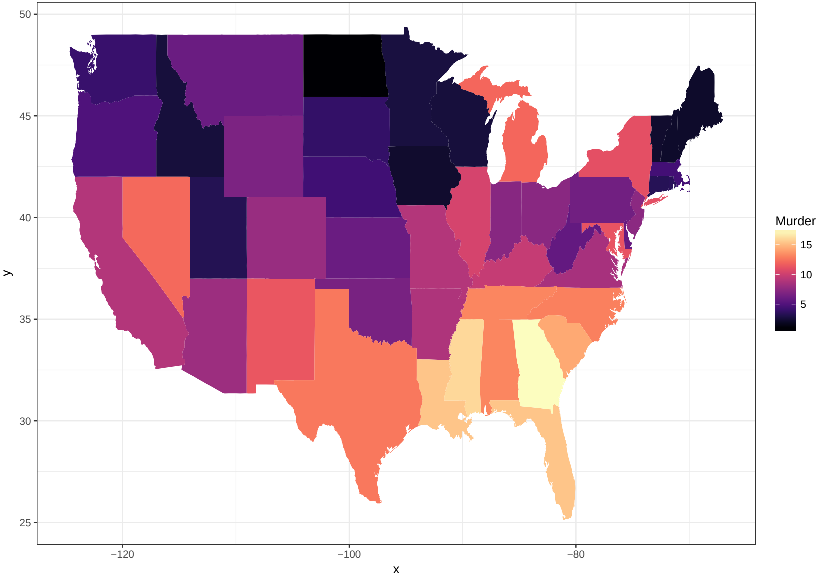
ggedit ★ Star

ggedit is aimed to interactively edit ggplot layers, scales and themes aesthetics

ggforce ★ Star

ggforce is aimed at providing missing functionality to ggplot2 through the extension

Literally so many things you can do...



very much inspiried by this example

Code for the previous example

```
library(maps)
library(viridis)
crimes <- data.frame(state = tolower(rownames(USArrests)), USArrests)
crimesm <- reshape2::melt(crimes, id = 1)

states_map <- map_data("state")
ggplot(crimes, aes(map_id = state)) +
    geom_map(aes(fill = Murder), map = states_map) +
    expand_limits(x = states_map$long, y = states_map$lat) +
scale_fill_viridis(option="magma") + theme_bw()
```

Another good tutorial

I was going to follow this tutorial, but then instead I decided to be more creative... Here it is anyway!!

An Introduction on How to Make Beautiful Charts With R and ggplot2
By Max Woolfe