Visualizing Data

Pedram Navid

August 14, 2016

Overview

Why Viz First?

Philosophy: start with the stuff you'll want

- so you'll stick around for the stuff you hate.

Visual Exploration of Data

- Most important part of data exploration
- Ability to see trends and relationships is unmatched by any statistical summary
- One of the most important methods of communication
- Analysis often hinges on proper visual exploration

Plotting Packages in R

- Lots of packages: base, lattice, ggplot2, others..
- In keeping with the philosophy of this tutorial, we will focus on ggplot2
- Two main functions: `ggplot()` and `qplot()`

```r
# Install if you haven't
# install.packages('ggplot2')
library(ggplot2)
```

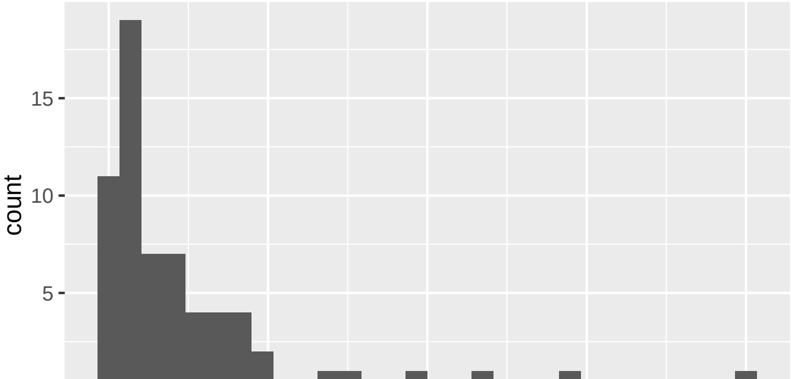
qplot()

qplot: fast plotting

- qplot gives you fast plotting, at the expensive of customizability
- great for throwing ideas out quickly and exploring new possibilities

```r
library(MASS)

# Histogram is default for one variable
qplot(Claims, data = Insurance)
```

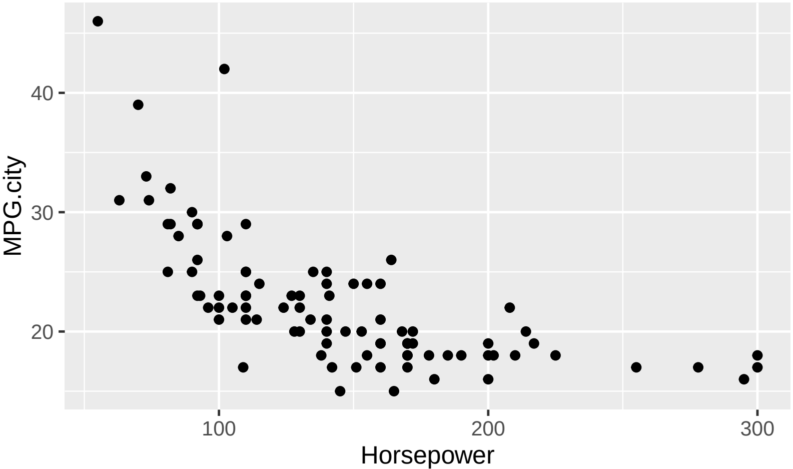


ggplot()

ggplot primer

Syntax is confusing at first.

```
ggplot(data=Cars93, aes(x = Horsepower, y = MPG.city)) +
  geom_point()
```



ggplot: aesthethics

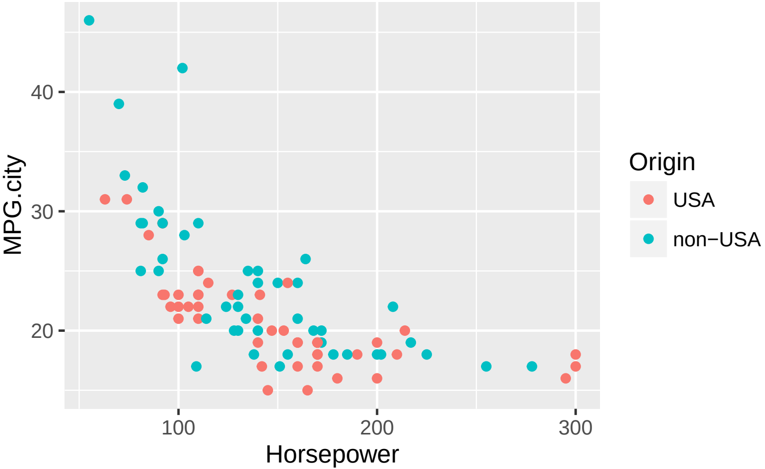
aesthethics

- aesthethics are something that the plot draws that varies with data
- examples:
  - colour of a point or line
  - size of a point, or line
  - fill of a bar, histogram
  - shape of a point
- use aes() to define them, either for the whole plot: `ggplot(data = bla, aes(x, y))`
- or for individual layers, if different layers have different aesthethics
  - `geom_line(aes(date, value1, colour = group)) + geom_line(aes(date, value2, colour = group))`

ggplot: aesthethics – examples

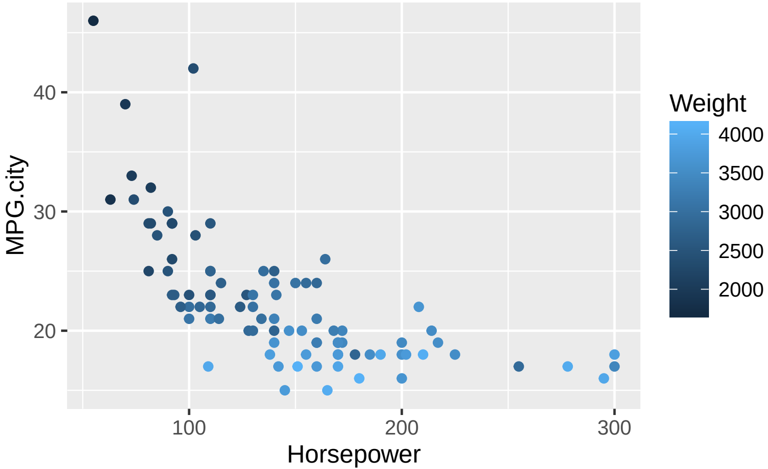
Colour

```
# Position (required) + discrete colour aesthethic
ggplot(data=Cars93, aes(x = Horsepower, y = MPG.city, colour = Origin)) +
  geom_point()
```



Continous colour

```
ggplot(data=Cars93, aes(Horsepower, MPG.city, colour = Weight)) +
  geom_point()
```



Size

```
ggplot(data=Cars93, aes(Horsepower, MPG.city, size = Weight)) +
```

ggplot: geoms

Overview of geoms

A geom is a thing that ggplot draws based on data. It will manipulate the data in some way (sometimes) and then draw it on a plot.

The ggplot2 cheatsheet is really helpful here: `https://www.rstudio.com/wp-content/uploads/2015/12/ggplot2-cheatsheet-2.0.pdf`

transformations

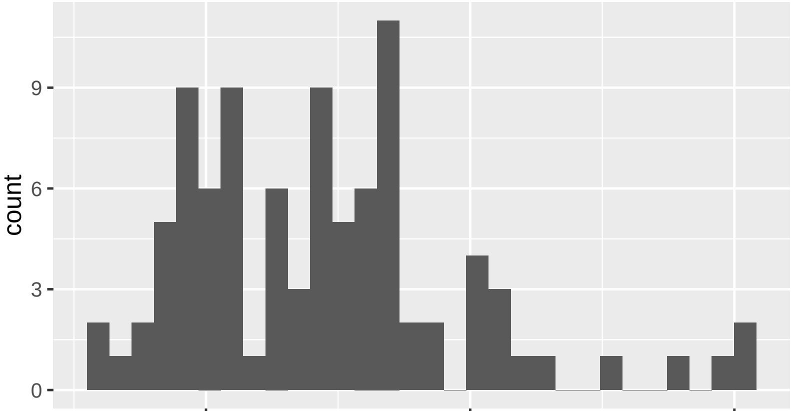
geoms may transform your data under the hood as needed.

ggplot geoms: 1 variable (continous)

histogram

```r
# If you don't provide bins, ggplot2 will (rightly) complain
ggplot(Cars93, aes(Horsepower)) +
  geom_histogram()
```

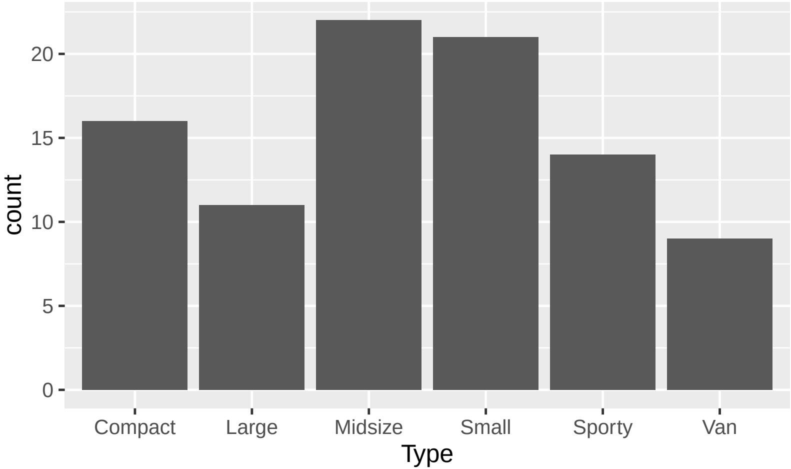
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



ggplot geoms: 1 variable (discrete)

bar charts

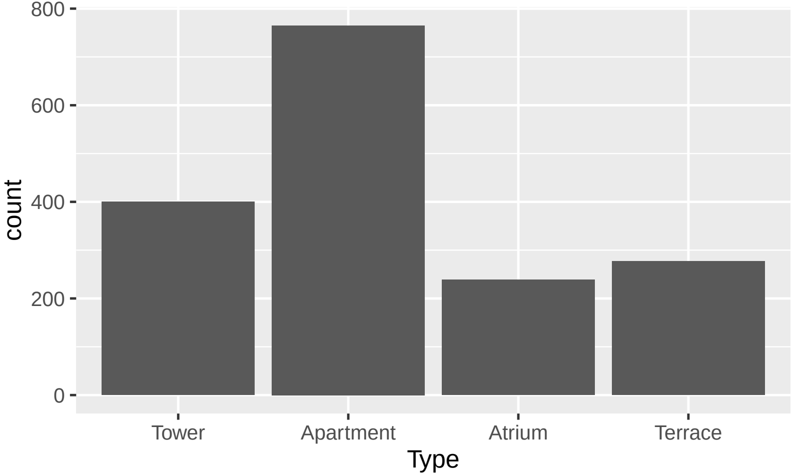
```
ggplot(Cars93, aes(Type)) +
  geom_bar()
```



ggplot geoms: 2 variables

bar chart – weighted

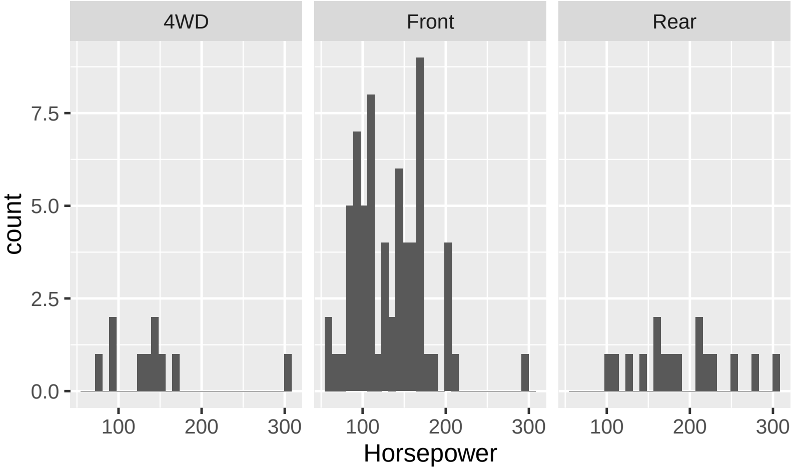
```
ggplot(housing, aes(Type, weight = Freq)) +
  geom_bar()
```



ggplot - facets

facet_wrap

```
ggplot(Cars93, aes(Horsepower)) +
  geom_histogram() +
  facet_wrap(~ DriveTrain)
```



Your Turn