Agenda Day 1

1. Announcement:
   How to make sure you have the right code from group work

2. Brief Review

3. Activity

4. Lecture: another graph for understanding distributions

5. Don't forget: homework due Thursday!

What have we learned?

long vs. wide

Wide data looks like this:

State	1960	1970	1980	1990	2000
New York	2	5	2	5	4
New Jersey	3	1	4	1	5
Arizona	3	9	8	7	5

While **long data** looks like this:

State	Year	Value
New York	1960	2
New York	1970	5
New York	1980	2
New York	1990	5
New York	2000	4
New Jersey	1960	3
New Jersey	1970	1
New Jersey	1980	4
…	…	…

read data + examine

```
head(flights)
```

year <int>	month <int>	day <int>	dep_time <int>	dep_delay <int>
2014	1	1	914	14
2014	1	1	1157	-3
2014	1	1	1902	2
2014	1	1	722	-8
2014	1	1	1347	2
2014	1	1	1824	4

6 rows | 1-10 of 17 columns

```
head()
str()
dim()
length()
```
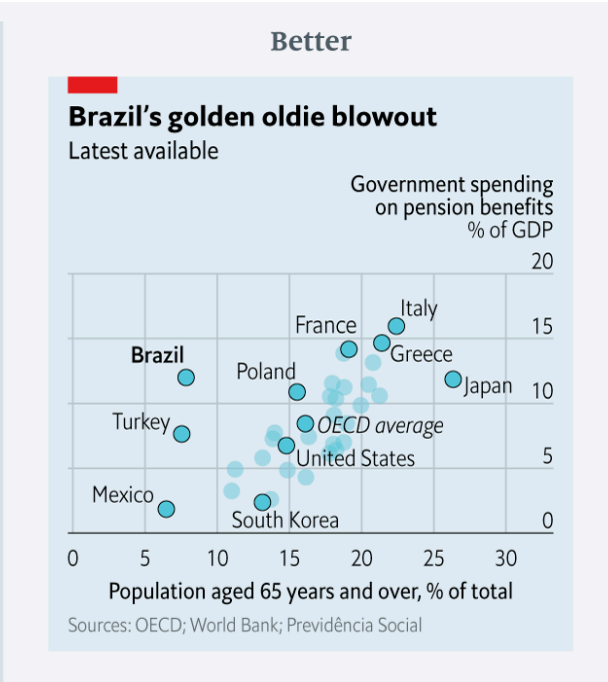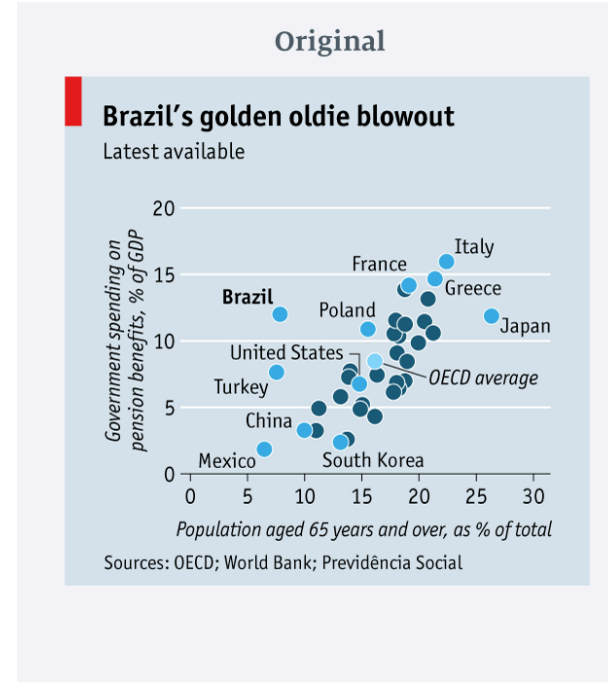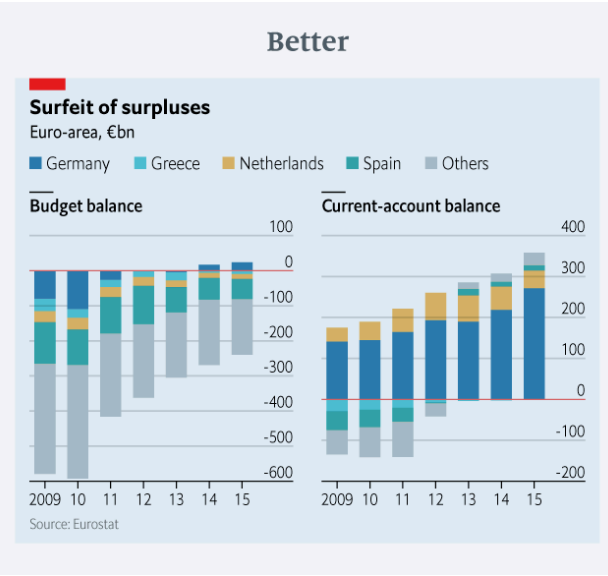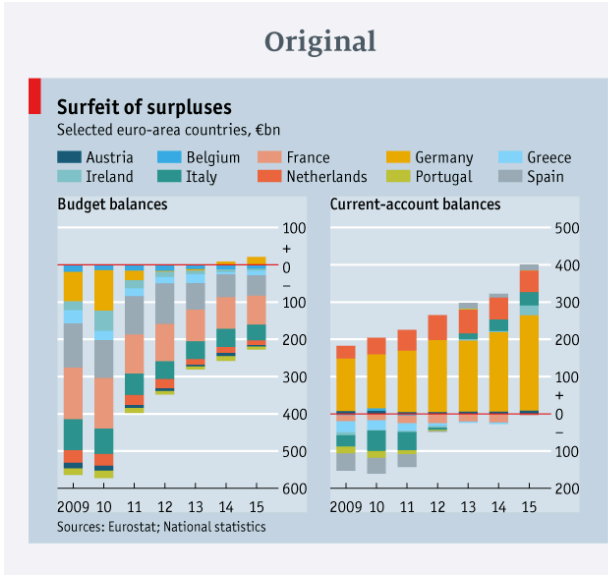
summarize data + manipulate

```
summary()
is.numeric()
min()
max()
median()
data.table()
group_by()
summarize()
```
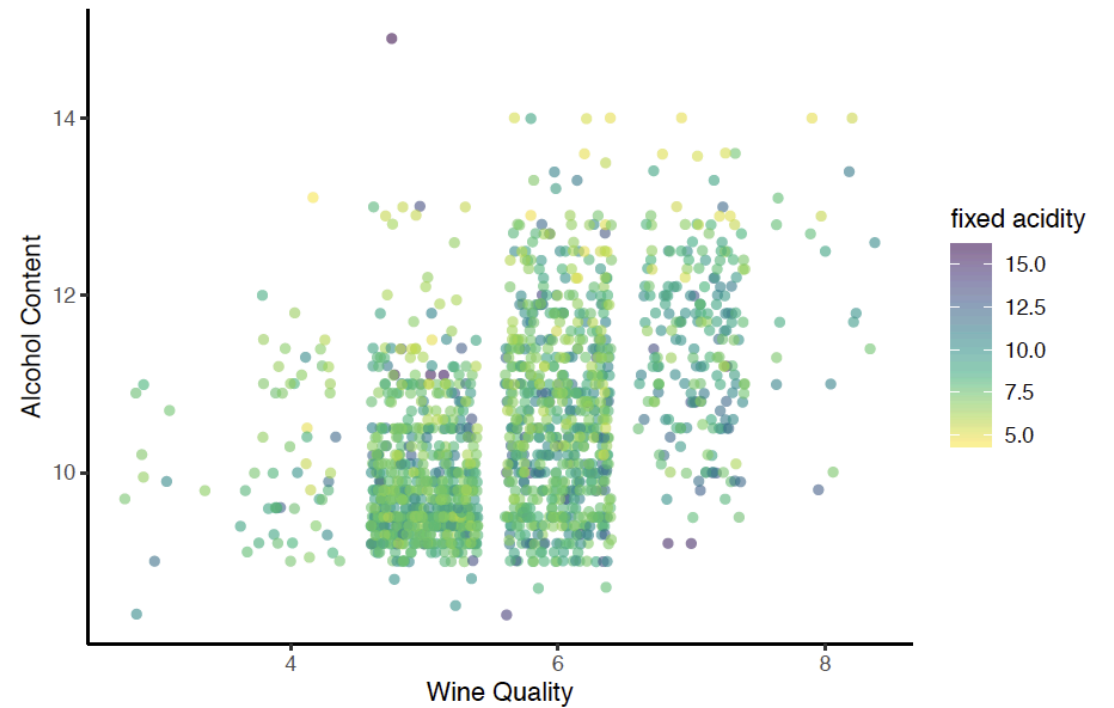
grammar of graphics & ggplot2
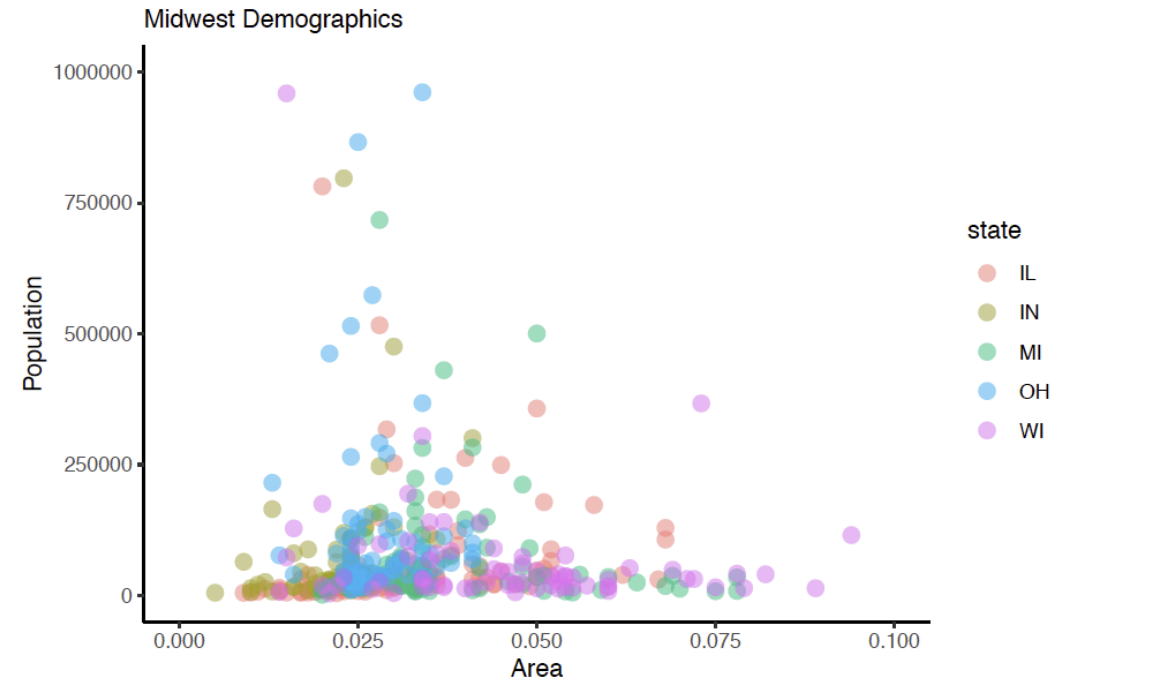
```
ggplot()
geom_plot()
geom_jitter()
theme()
labs()
scale_x()
```
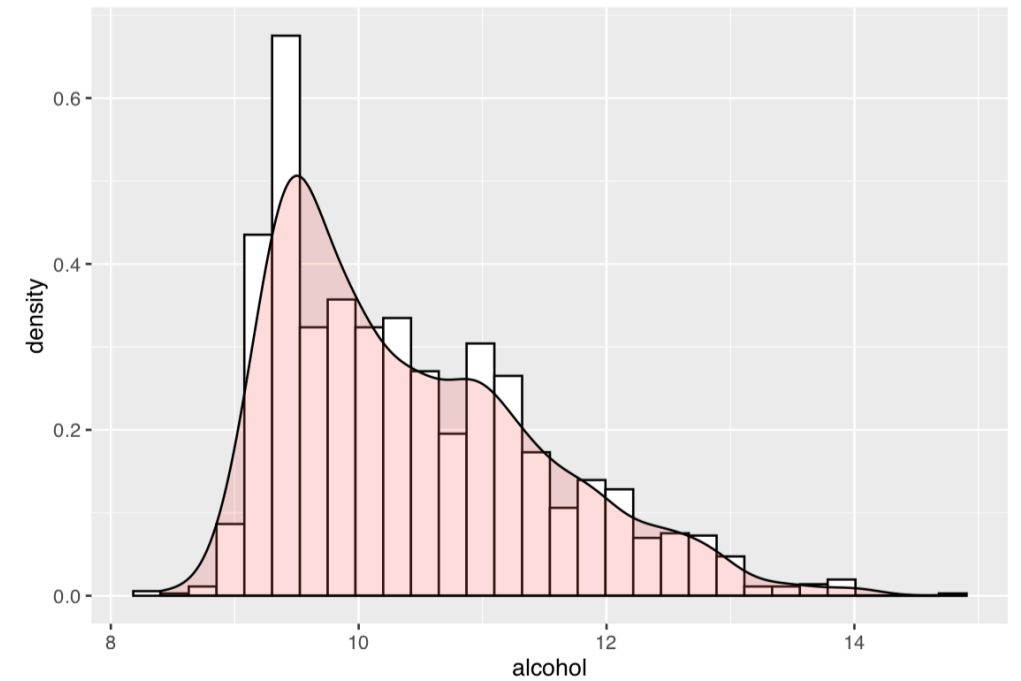
golden rules

Surfeit of surpluses
Selected euro-area countries, €bn

Austria Belgium France Germany Greece
Ireland Italy Netherlands Portugal Spain

Budget balances

Current-account balances

2009 10 11 12 13 14 15

Sources: Eurostat; National statistics

Better

Surfeit of surpluses
Euro-area, €bn

Germany Greece Netherlands Spain Others

Budget balance

Current-account balance

2009 10 11 12 13 14 15

Source: Eurostat

Original

Brazil's golden oldie blowout
Latest available

Government spending on pension benefits, % of GDP

Italy
France
Brazil
Greece
Poland
Japan
United States
OECD average
Turkey
China
Mexico
South Korea

Population aged 65 years and over, as % of total

Sources: OECD; World Bank; Previdência Social

Better

Brazil's golden oldie blowout
Latest available

Government spending on pension benefits
% of GDP

Italy
France
Brazil
Greece
Poland
Japan
OECD average
United States
Turkey
Mexico
South Korea

Population aged 65 years and over, % of total

Sources: OECD; World Bank; Previdência Social

scatterplots

Midwest Demographics
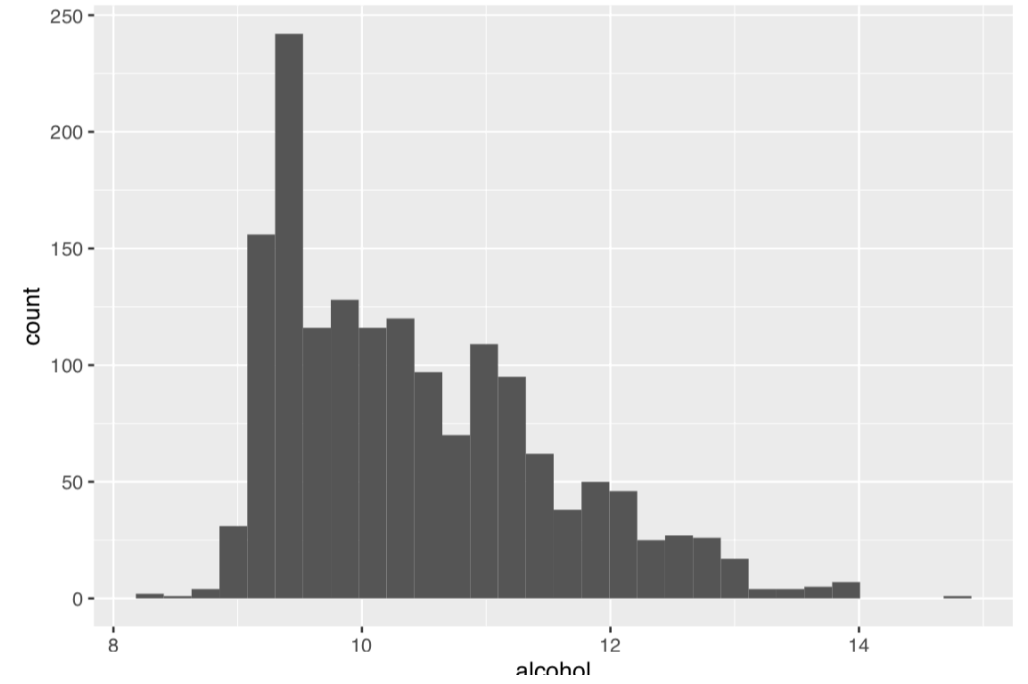
distributions

DATA QUICKFIRE

DATA QUICKFIRE

- Data: metro-nash-schools.csv
- Goal: one beautiful plot
- Must be: a scatterplot, a histogram, or a density plot
- Must be: done in Rstudio
- Time: 30 minutes
- If 'tapped' you will slack me the graphic for display
- Winner: gets a prize