Uncertainty

Session 6

PMAP 8921: Data Visualization with R
Andrew Young School of Policy Studies
Summer 2023

Plan for today

Communicating uncertainty

Visualizing uncertainty

Communicating uncertainty

The Bay of Pigs

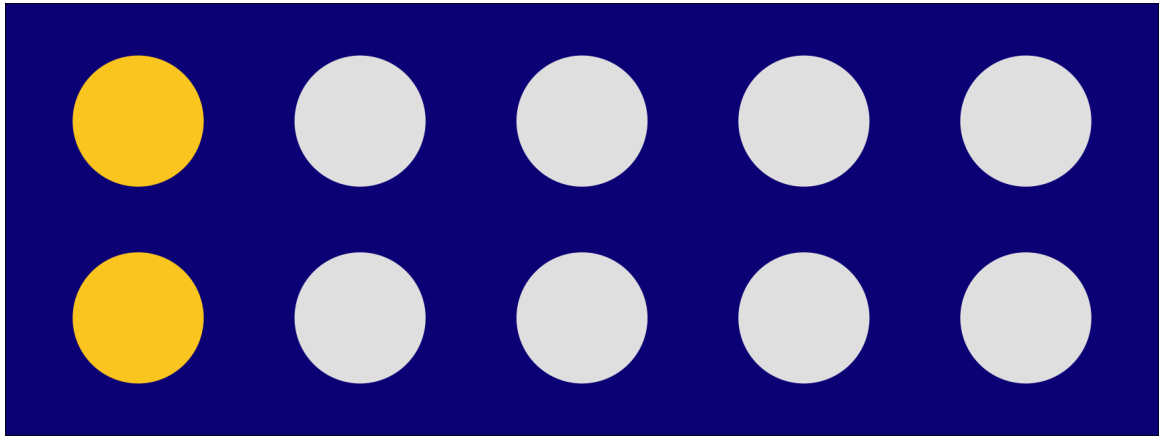
Joint Chiefs said "fair chance of success"

In Pentagon-speak, that meant 3:1 odds of failure

25% chance of success!

Misperceptions of probability

1 in 5 vs. 20%



Misperceptions of probability

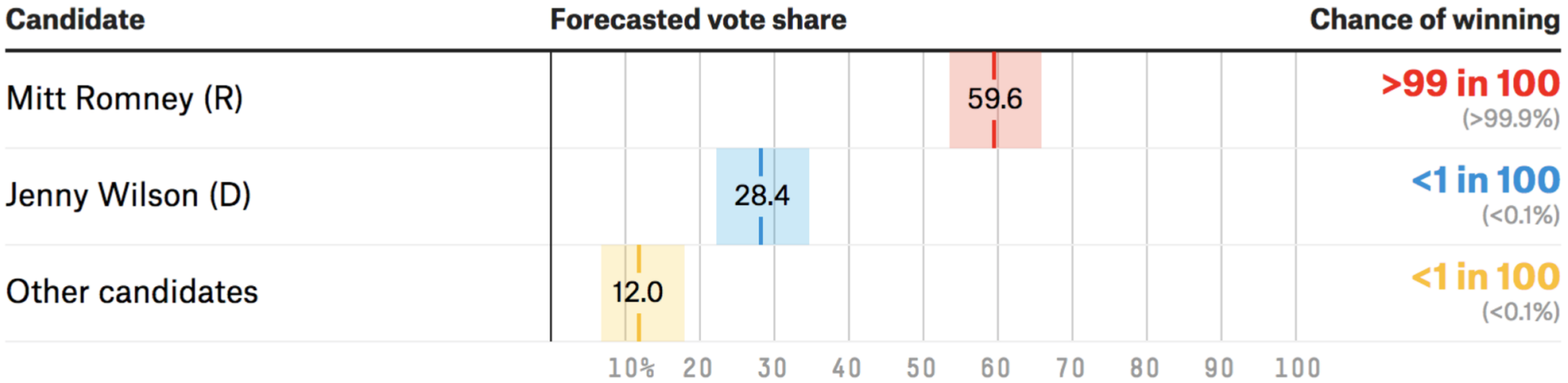
Utah SOLID R

<1 in 100

Chance the Democrat wins (<0.1%)

>99 in 100

Chance the Republican wins (>99.9%)

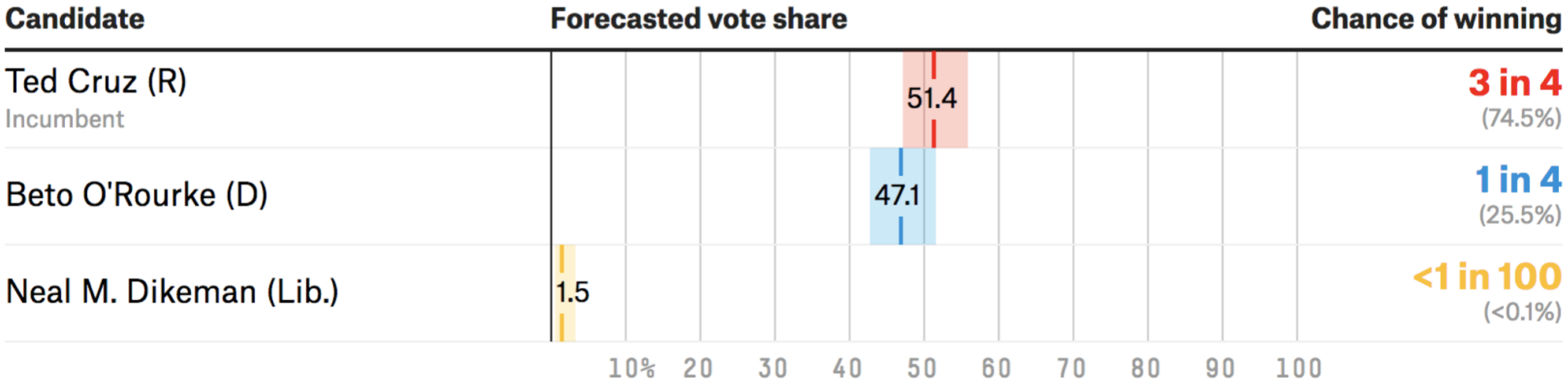


Misperceptions of probability

Texas LEAN R

1 in 4
Chance the Democrat wins (25.5%)

3 in 4
Chance the Republican wins (74.5%)



Misperceptions of probability

Chance of rain = Probability × Area

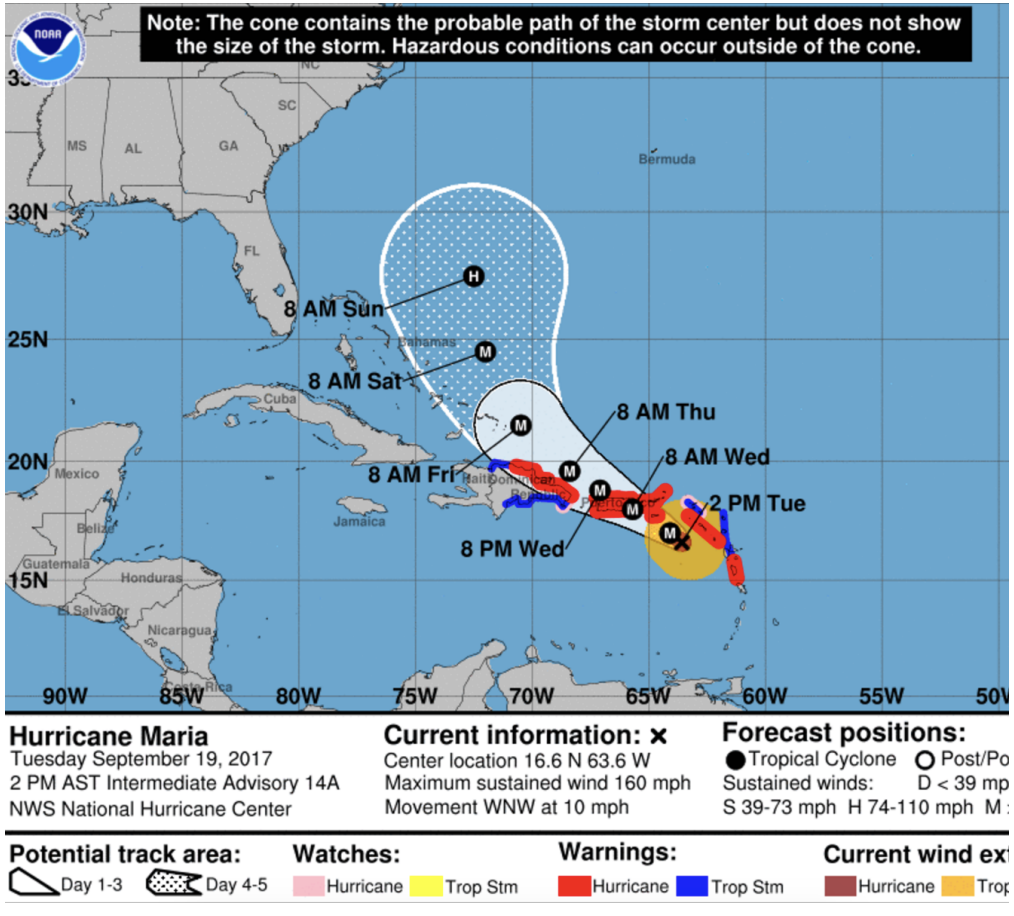
100% chance in 1/3 of the city

0% chance in 2/3 of the city

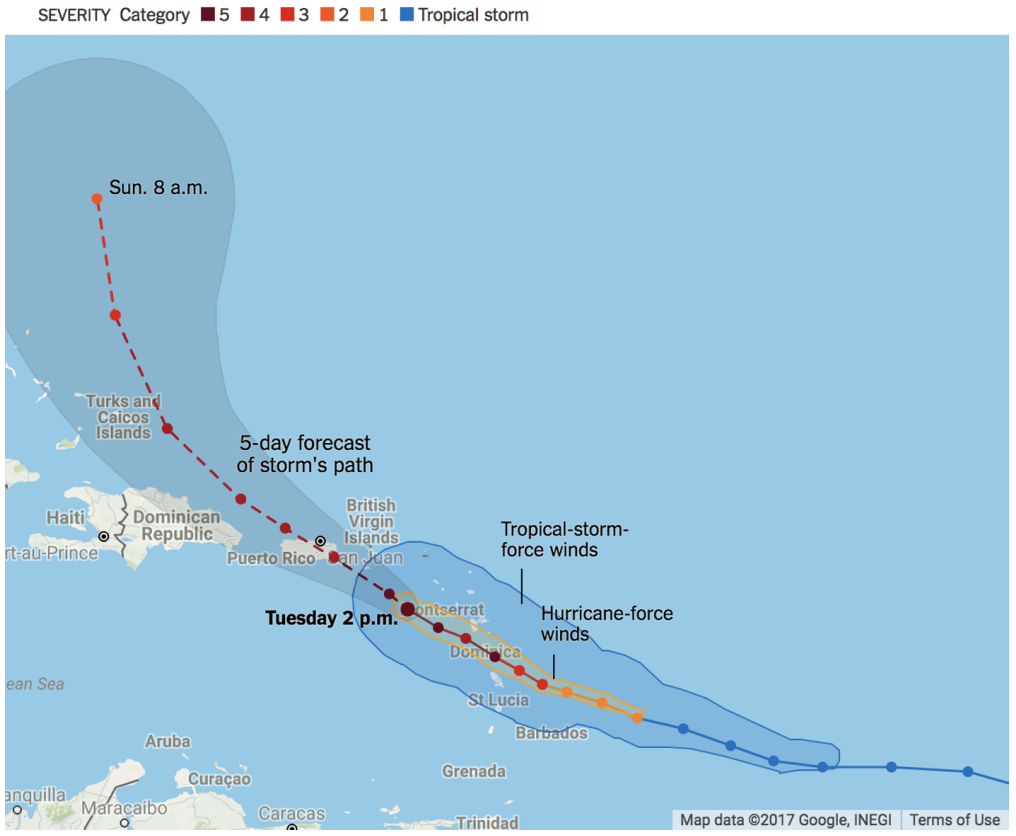
Chance of rain for city = 33%

Misperceptions of probability

Misperceptions of probability

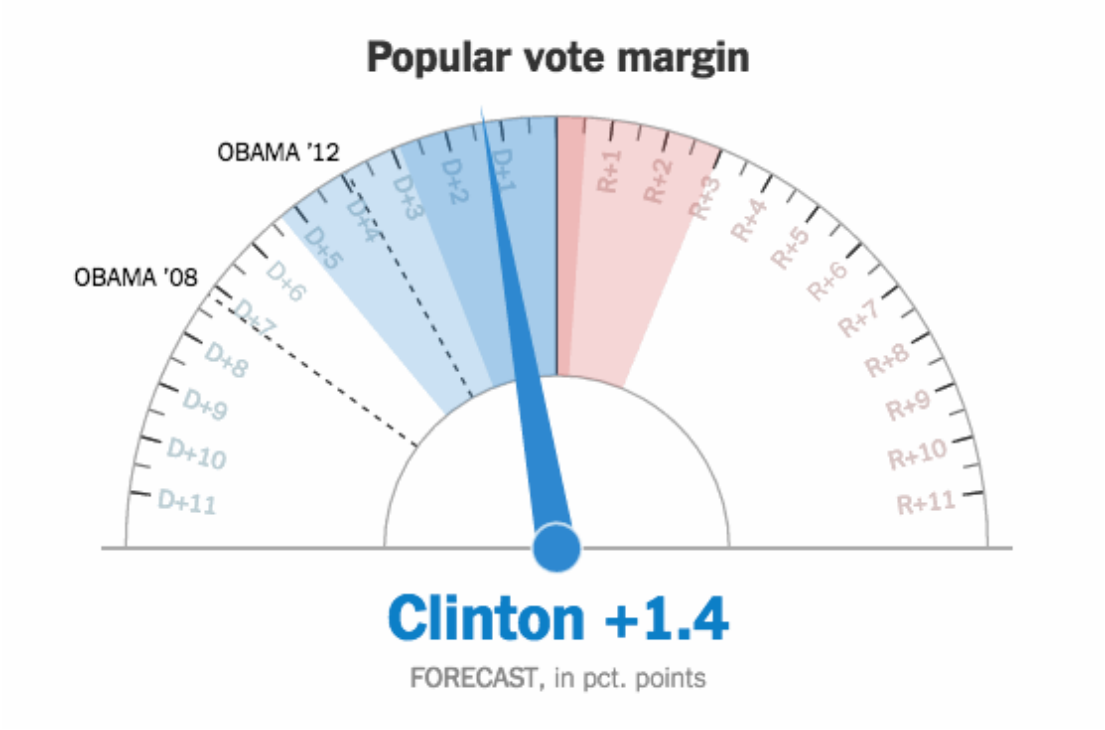


Hurricane Maria map, NOAA



Hurricane Maria map, New York Times

The needle

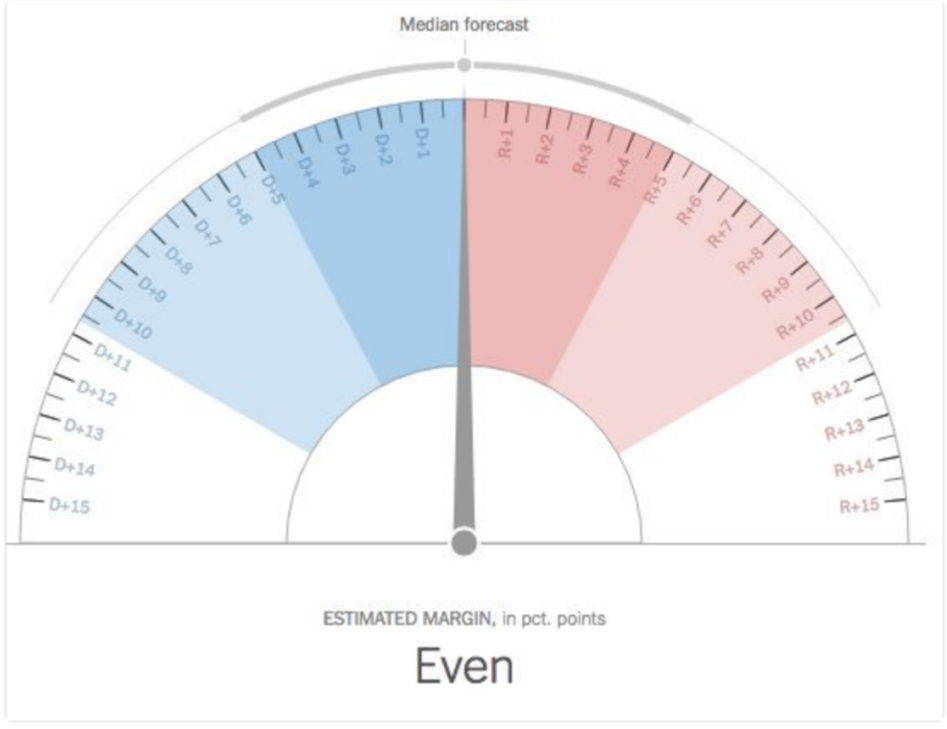

Popular vote margin
OBAMA '12
OBAMA '08
D+1
D+2
D+3
D+4
D+5
D+6
D+7
D+8
D+9
D+10
D+11
R+1
R+2
R+3
R+4
R+5
R+6
R+7
R+8
R+9
R+10
R+11
Clinton +1.4
FORECAST, in pct. points

The needle

Nate Cohn ✓
@Nate_Cohn

Following

The GA-6 live model is live.
nytimes.com/elections/resu …



4:47 PM - 20 Jun 2017 from Manhattan, NY

Virgil Texas ✓ @virgiltexas · Jun 20
Replying to @Nate_Cohn
Nate

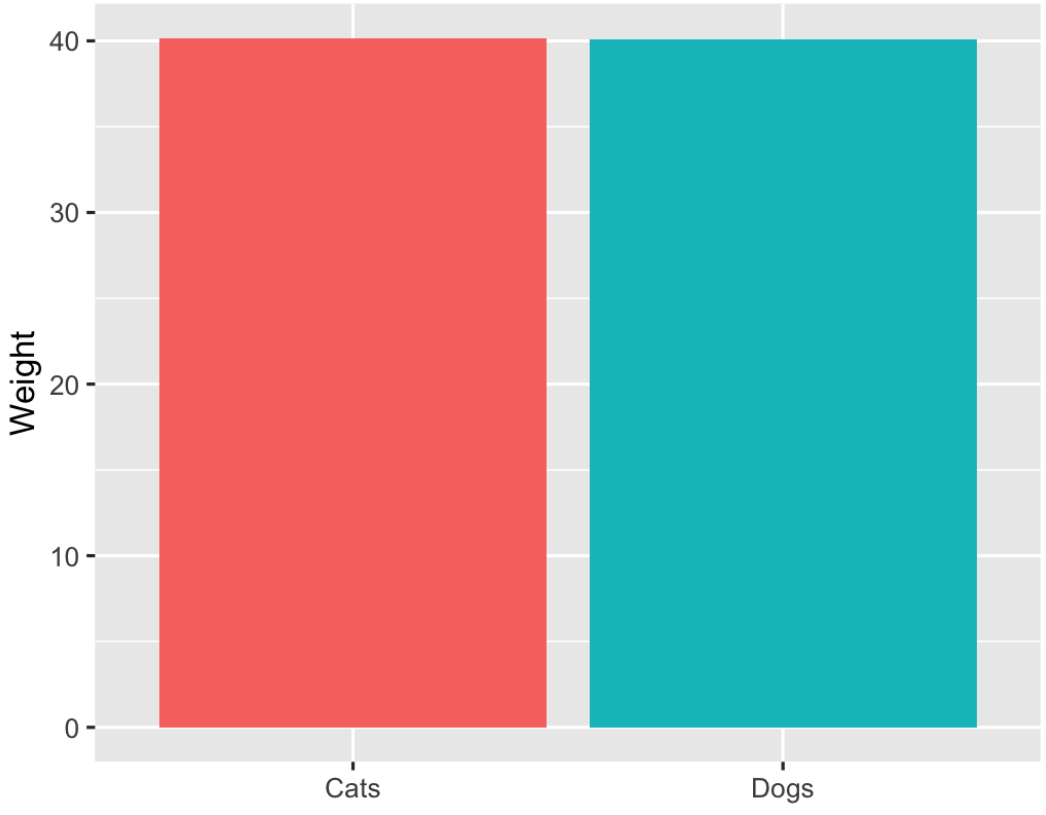
Alp Ozcelik ✓ @alplicable · Jun 20
Replying to @Nate_Cohn
DO NOT DO THIS TO ME AGAIN

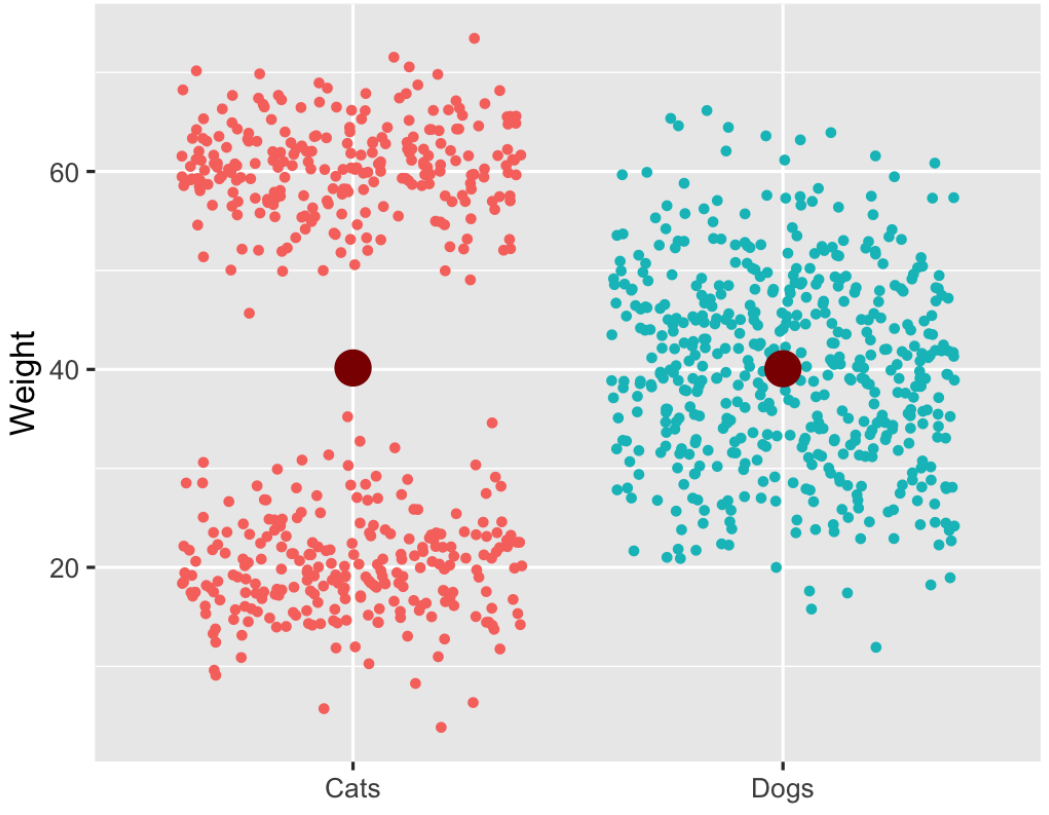
Sarcasmorator @Sarcasmorator · Jun 20
Replying to @Nate_Cohn @jacquicollins_
ah, yes, the election stress-o-meter

2 36

Visualizing uncertainty

Problems with single numbers





More information is always better

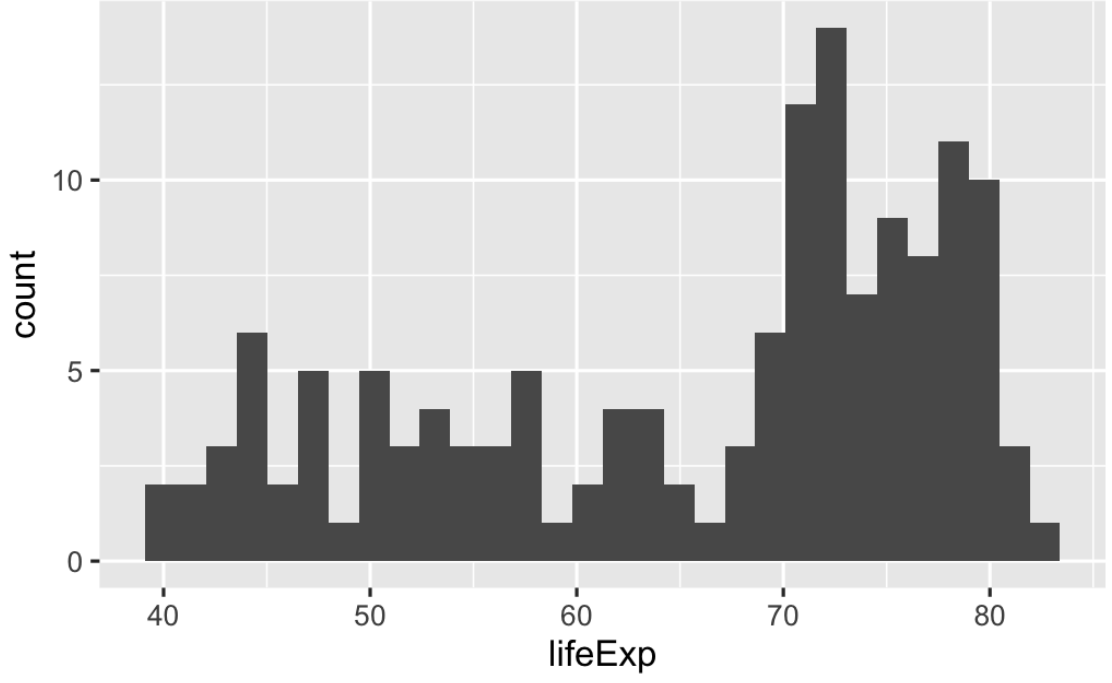
Avoid visualizing single numbers when you have a whole range or distribution of numbers

- Uncertainty in single variables
- Uncertainty across multiple variables
- Uncertainty in models and simulations

Histograms

Put data into equally spaced buckets (or bins), plot how many rows are in each bucket

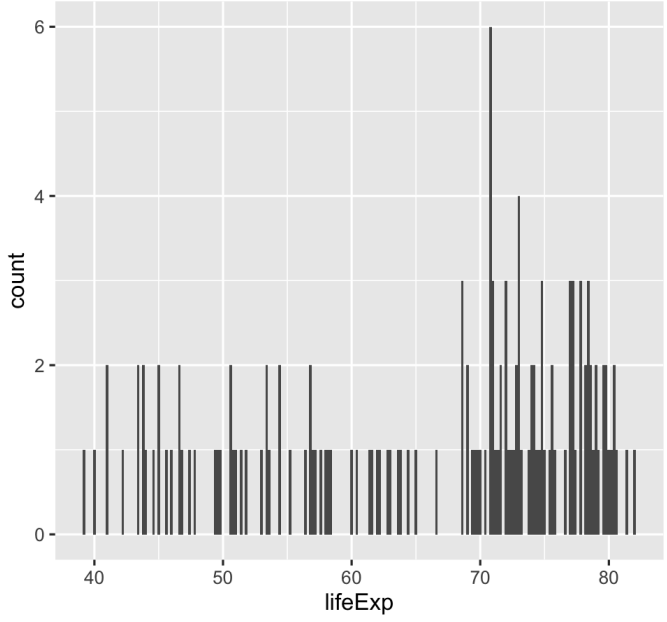
```
library(gapminder)

gapminder_2002 <- gapminder %>%
  filter(year == 2002)

ggplot(gapminder_2002,
       aes(x = lifeExp)) +
  geom_histogram()
```



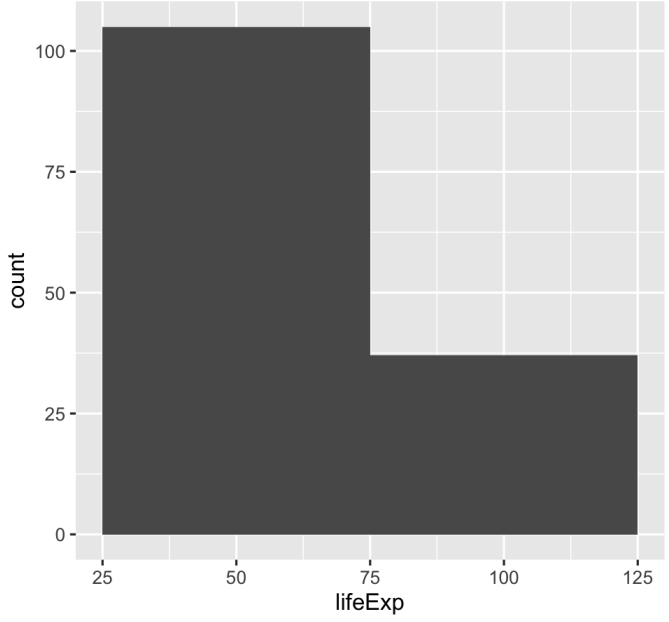
Histograms: Bin width

No official rule for what makes a good bin width

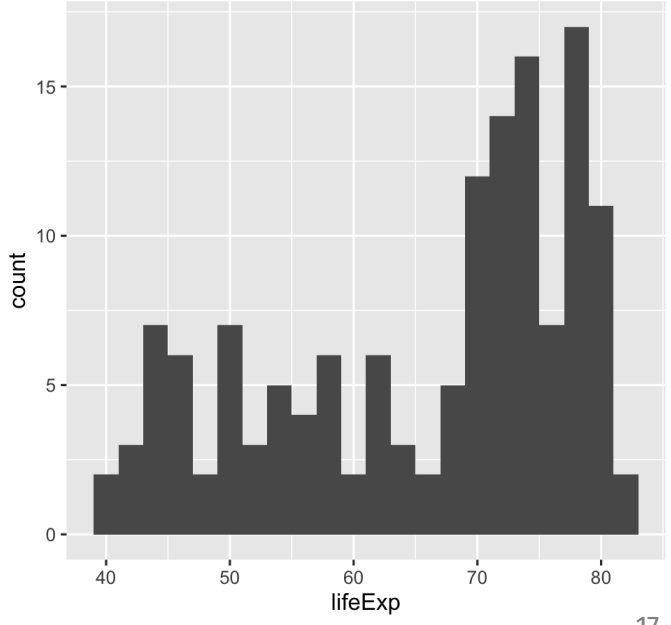
Too narrow:
`binwidth = 0.2`



Too wide:
`binwidth = 50`



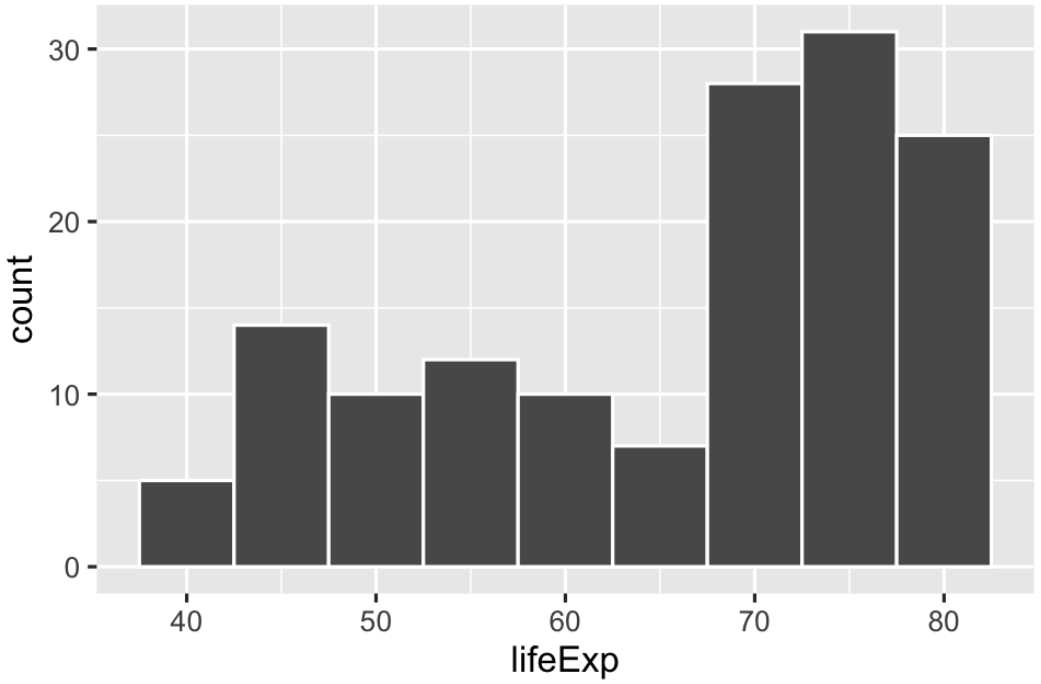
(One type of) just right:
`binwidth = 2`



Histogram tips

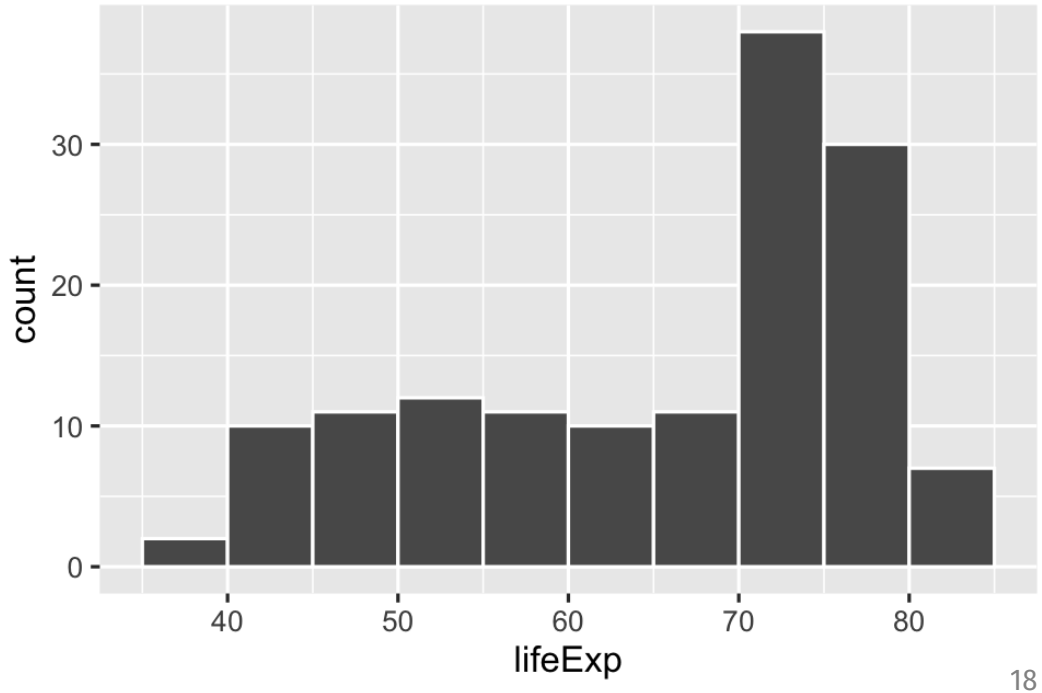
Add a border to the bars for readability

```
geom_histogram(..., color = "white")
```



Set the boundary; bucket now 50–55, not 47.5–52.5

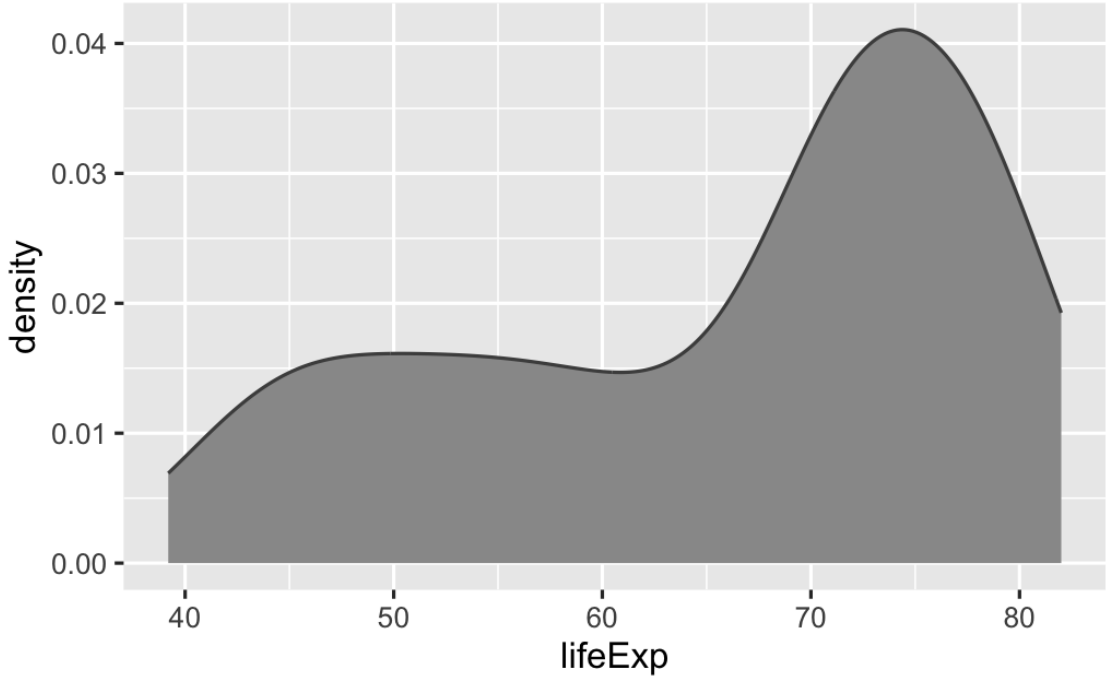
```
geom_histogram(..., boundary = 50)
```



Density plots

Use calculus to find the probability of each x value

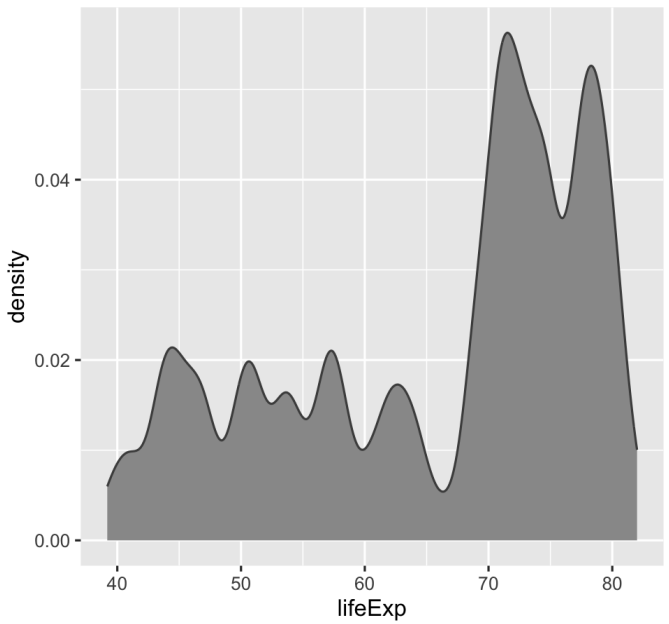
```
ggplot(gapminder_2002,
       aes(x = lifeExp)) +
  geom_density(fill = "grey60",
               color = "grey30")
```



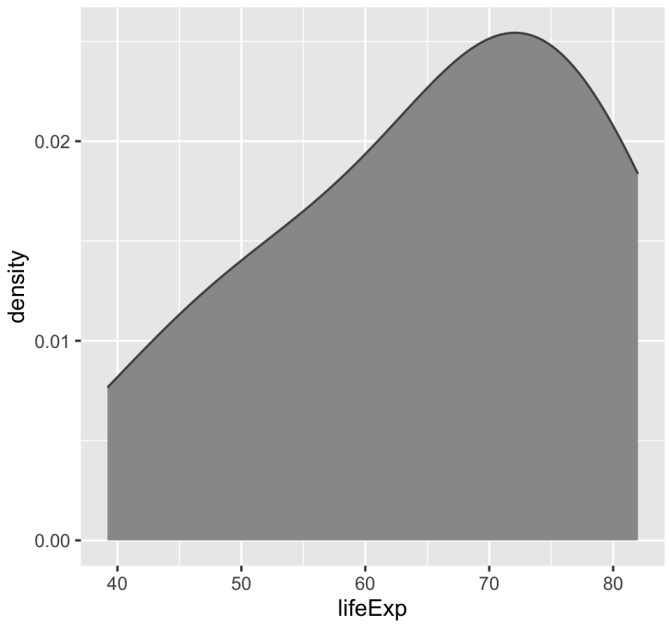
Density plots: Kernels and bandwidths

Different options for calculus change the plot shape

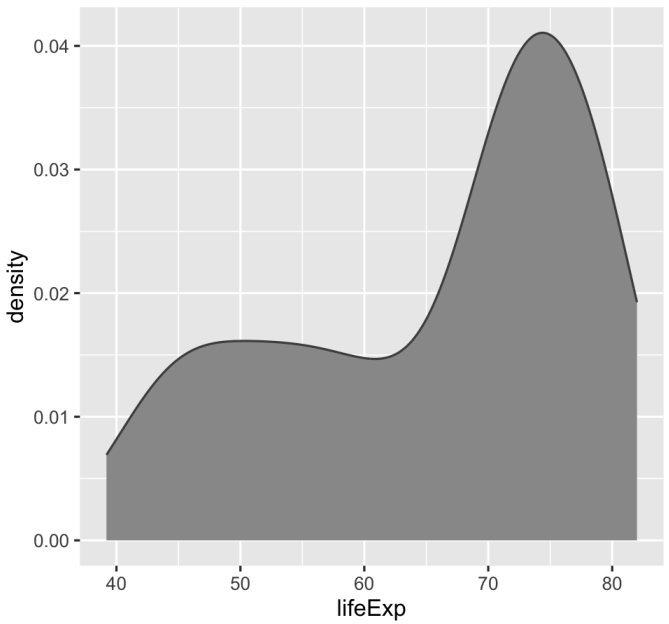
`bw = 1`



`bw = 10`



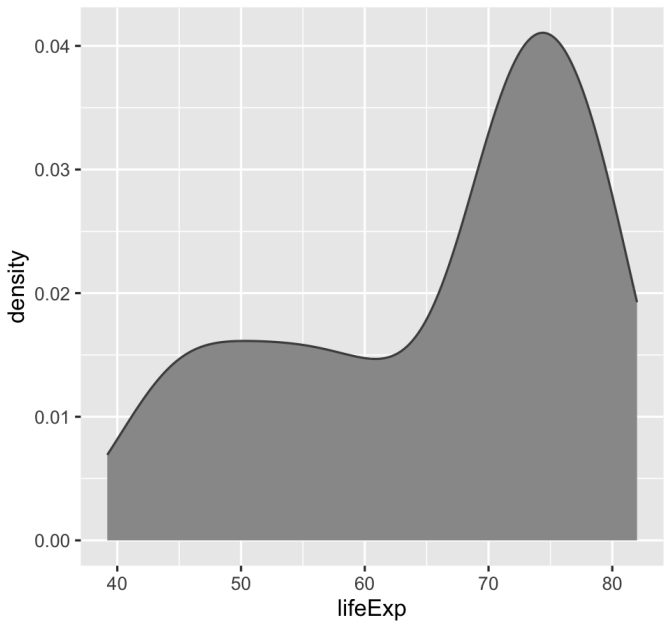
`bw = "nrd0"(default)`



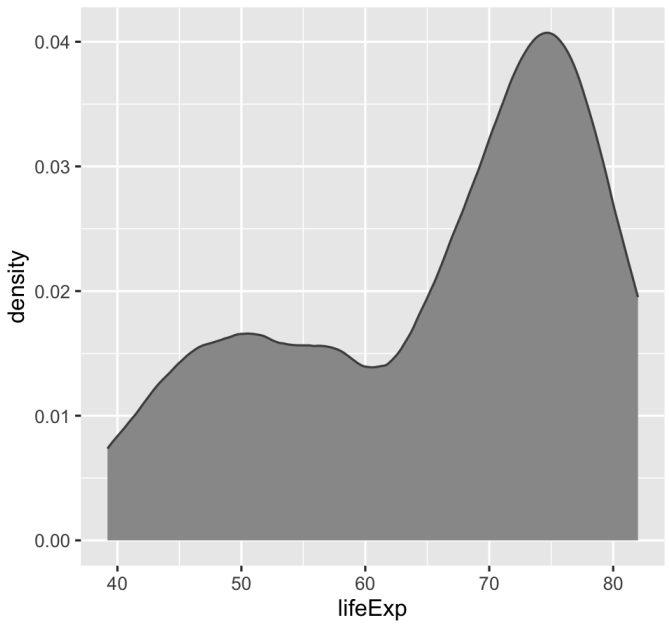
Density plots: Kernels and bandwidths

Different options for calculus change the plot shape

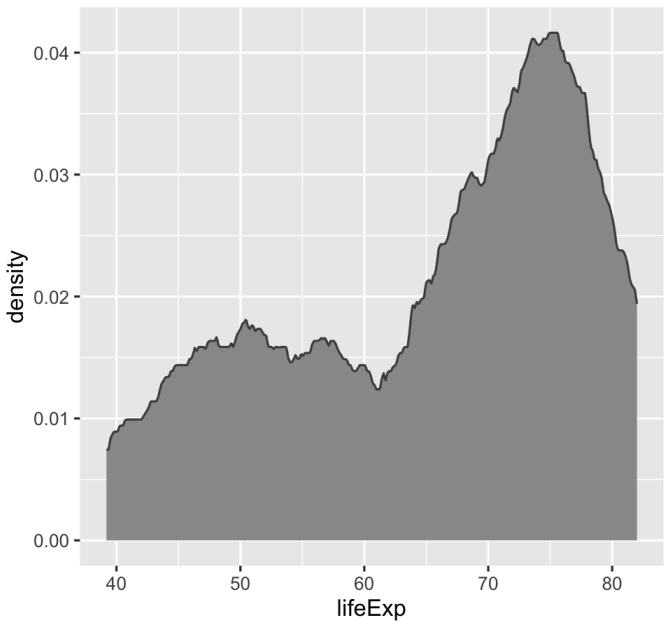
`kernel = "gaussian"`



`"epanechnikov"`



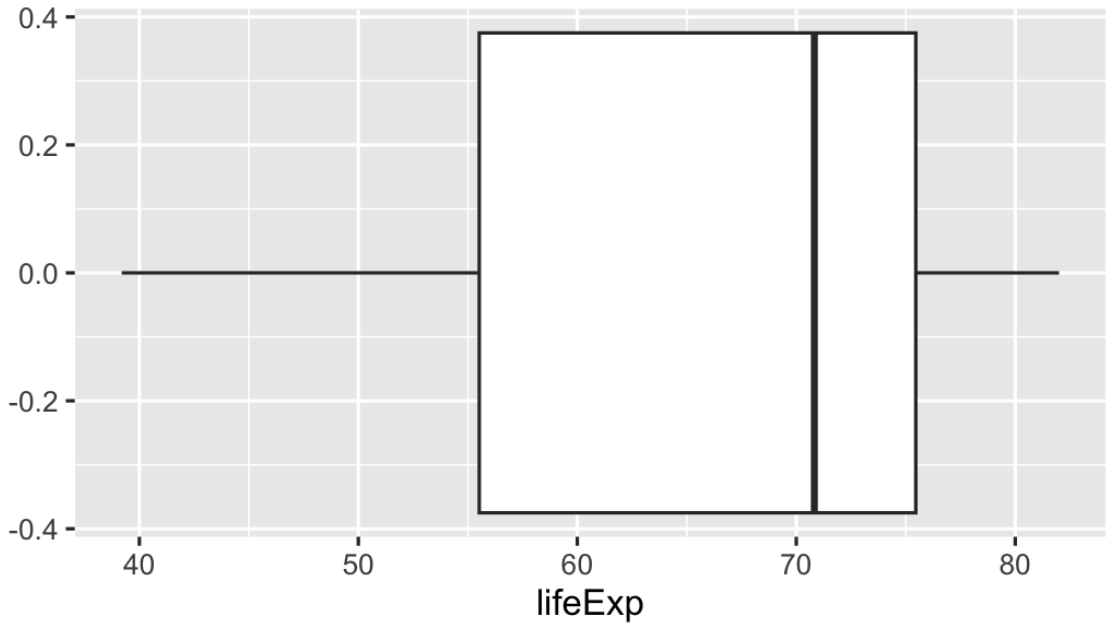
`"rectangular"`



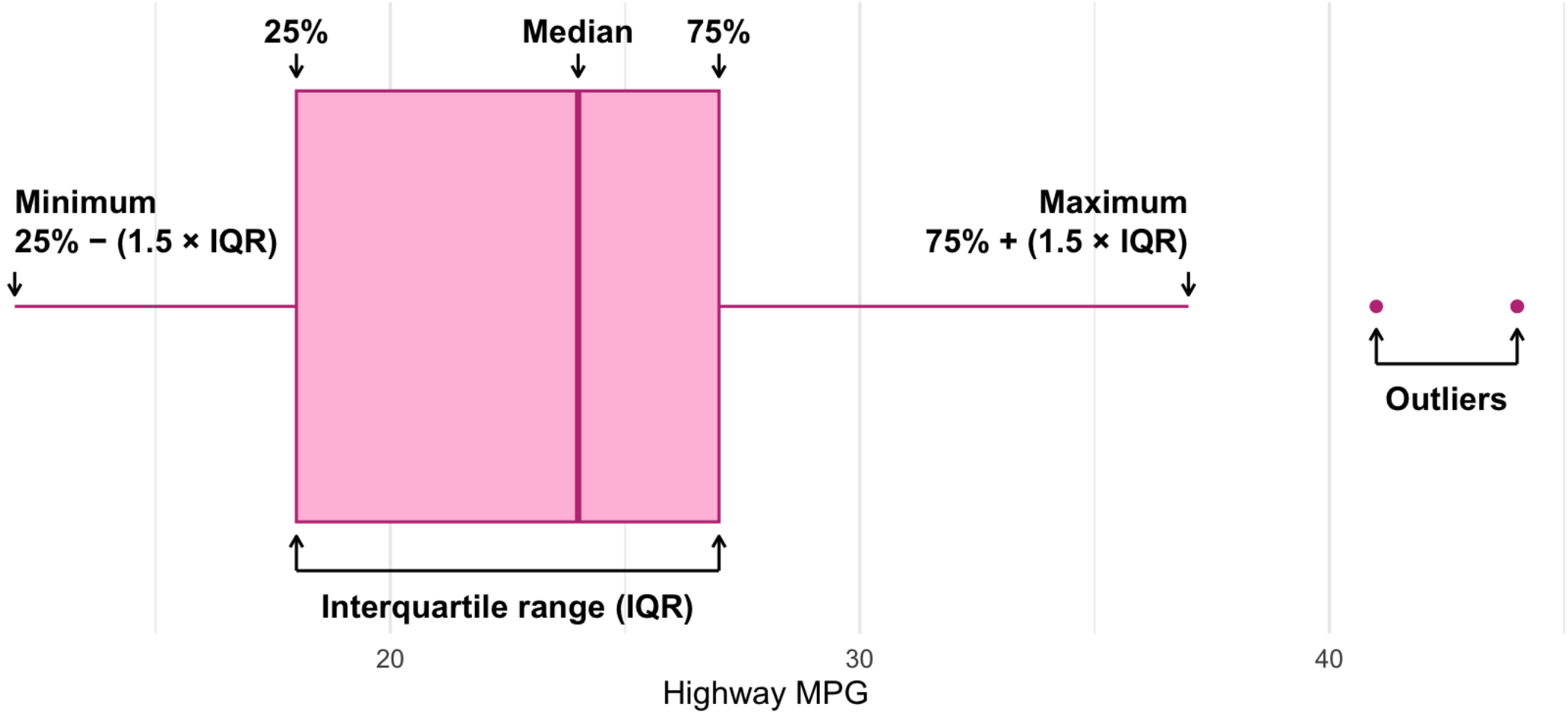
Box plots

Show specific distributional numbers

```
ggplot(gapminder_2002,
       aes(x = lifeExp)) +
  geom_boxplot()
```



Box plots

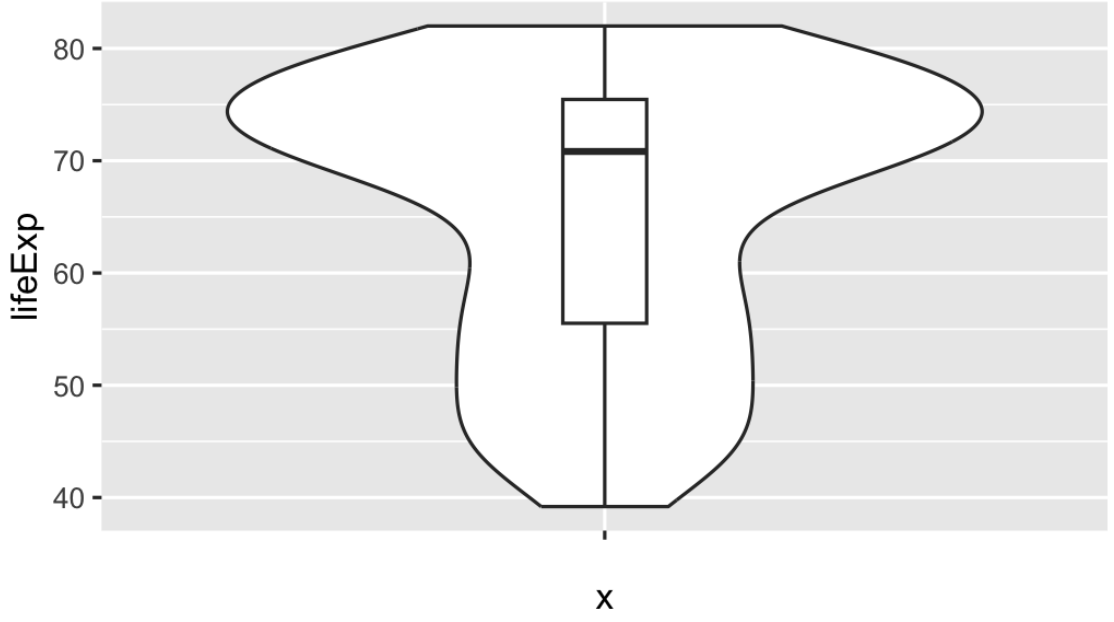

25%
Median
75%
Minimum
25% − (1.5 × IQR)
Maximum
75% + (1.5 × IQR)
Outliers
Interquartile range (IQR)
20
30
40
Highway MPG

Violin plots

Mirror density plot and flip

Often helpful to overlay other things on it

```
ggplot(gapminder_2002,
       aes(x = "",
           y = lifeExp)) +
  geom_violin() +
  geom_boxplot(width = 0.1)
```



Uncertainty across multiple variables

Visualize the distribution of a single variable across groups

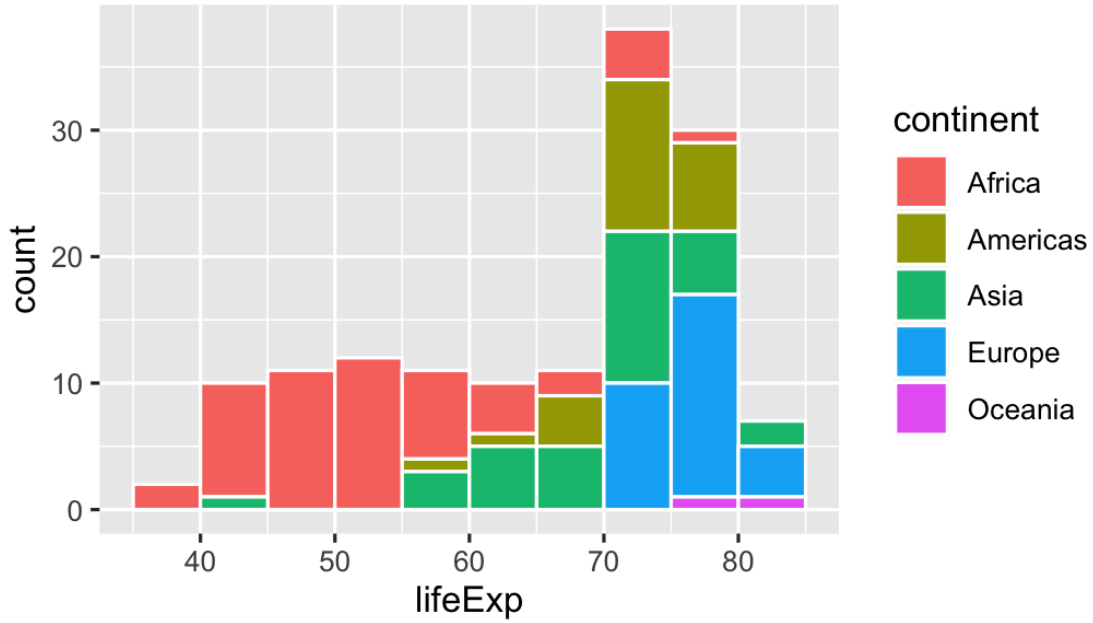
Add a `fill` aesthetic or use faceting!

Multiple histograms

Fill with a different variable

This is bad and really hard to read though

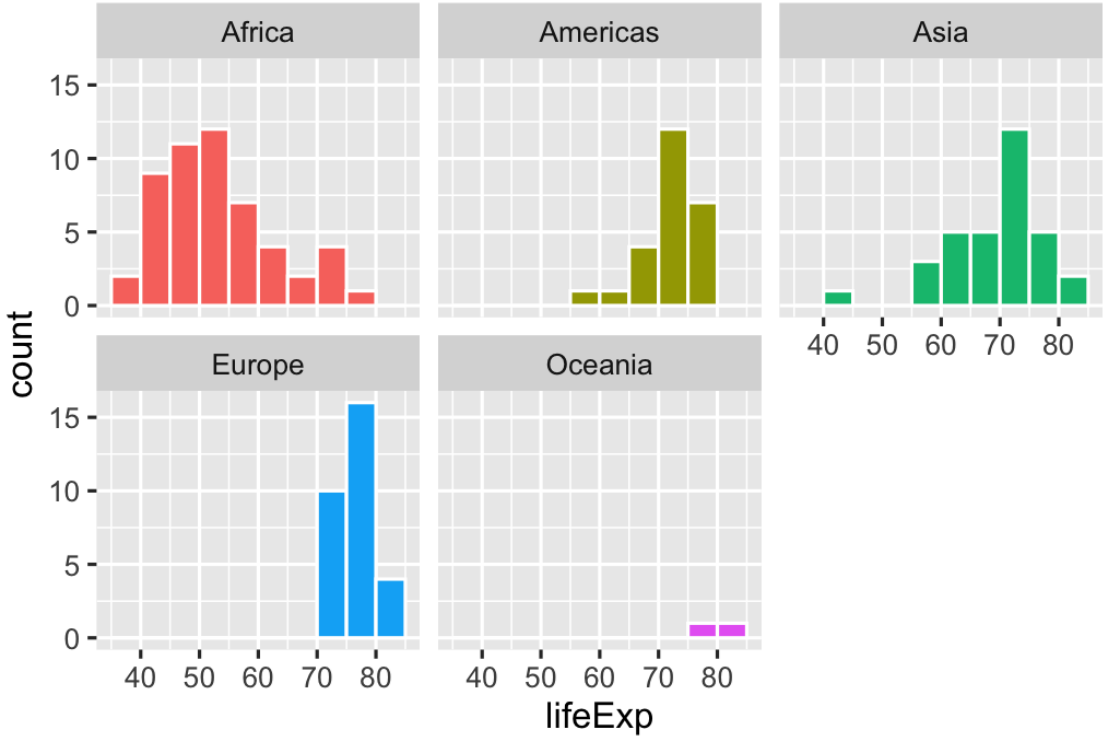
```
ggplot(gapminder_2002,
       aes(x = lifeExp,
           fill = continent)) +
  geom_histogram(binwidth = 5,
                 color = "white",
                 boundary = 50)
```



Multiple histograms

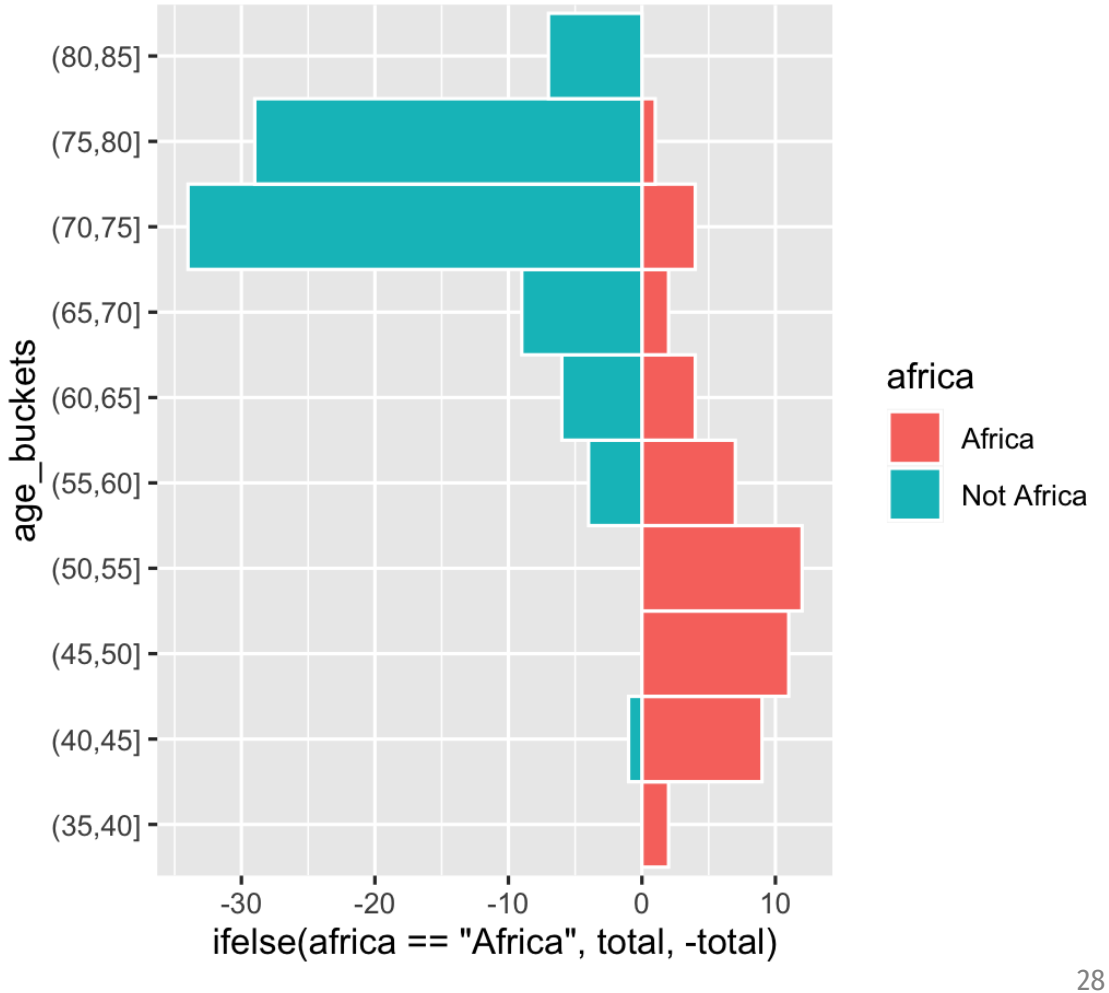
Facet with a different variable

```
ggplot(gapminder_2002,
       aes(x = lifeExp,
           fill = continent)) +
  geom_histogram(binwidth = 5,
                 color = "white",
                 boundary = 50) +
  guides(fill = "none") +
  facet_wrap(vars(continent))
```



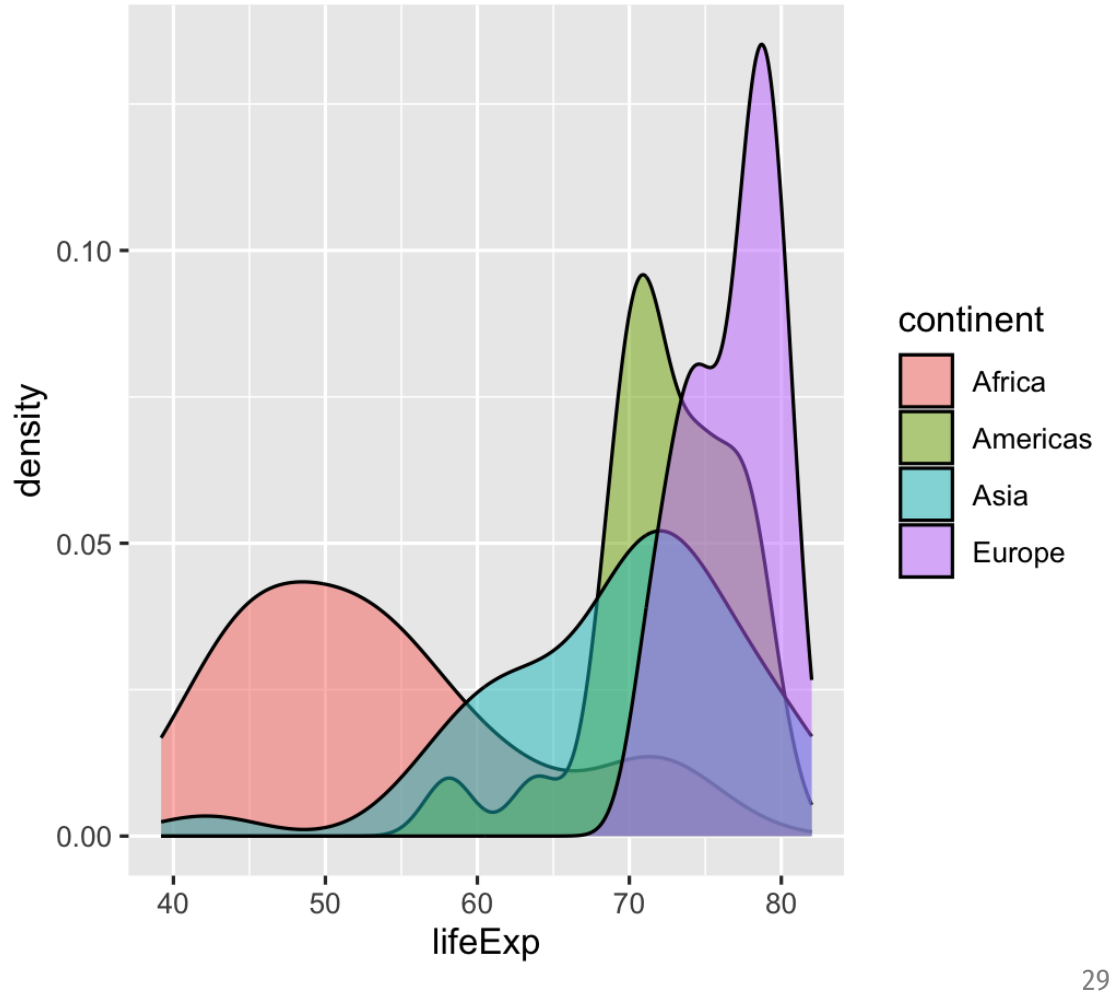
Pyramid histograms

```
gapminder_intervals <- gapminder %>%
  filter(year == 2002) %>%
  mutate(africa =
           ifelse(continent == "Africa",
                  "Africa",
                  "Not Africa")) %>%
  mutate(age_buckets =
           cut(lifeExp,
               breaks = seq(30, 90, by = 5))) %>%
  group_by(africa, age_buckets) %>%
  summarize(total = n())

ggplot(gapminder_intervals,
       aes(y = age_buckets,
           x = ifelse(africa == "Africa",
                      total, -total),
           fill = africa)) +
  geom_col(width = 1, color = "white")
```



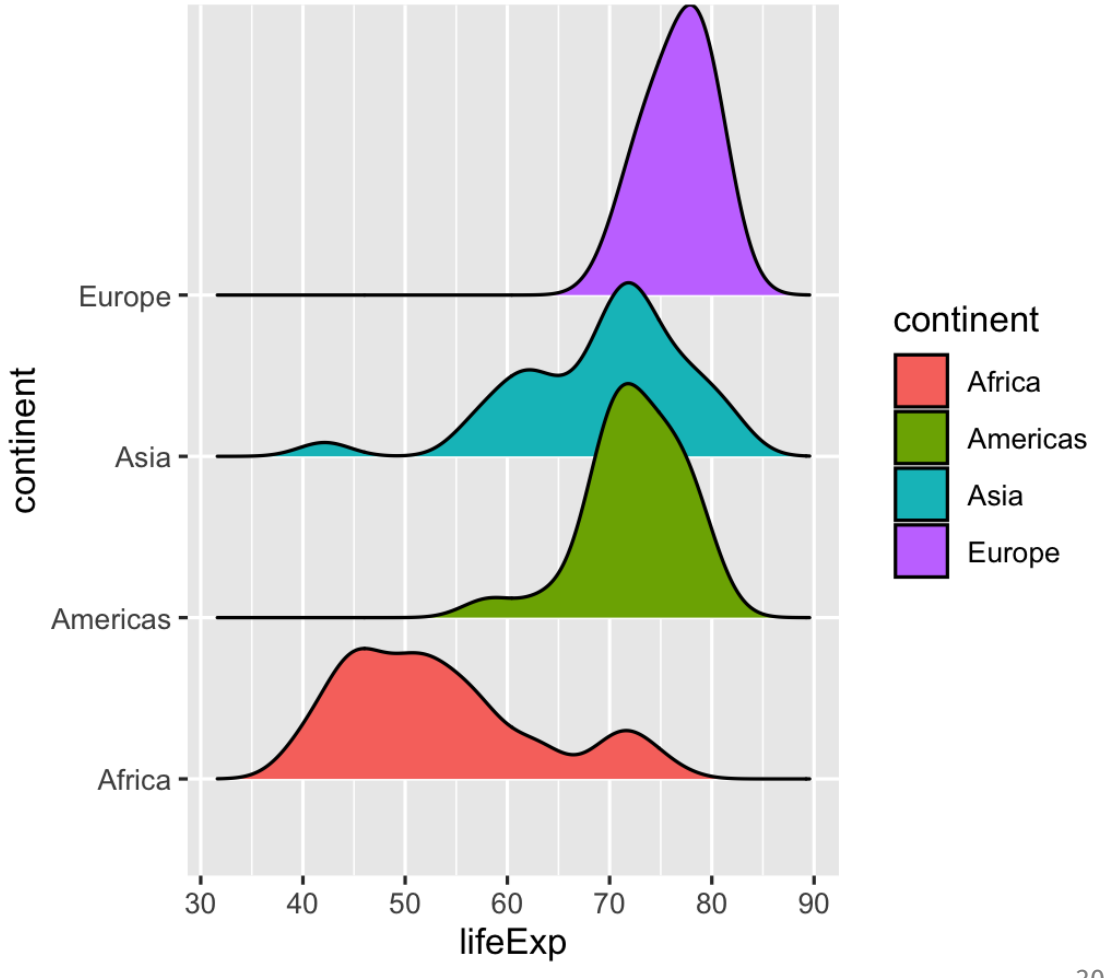
Multiple densities: Transparency

```
ggplot(filter(gapminder_2002,
              continent != "Oceania"),
       aes(x = lifeExp,
           fill = continent)) +
  geom_density(alpha = 0.5)
```

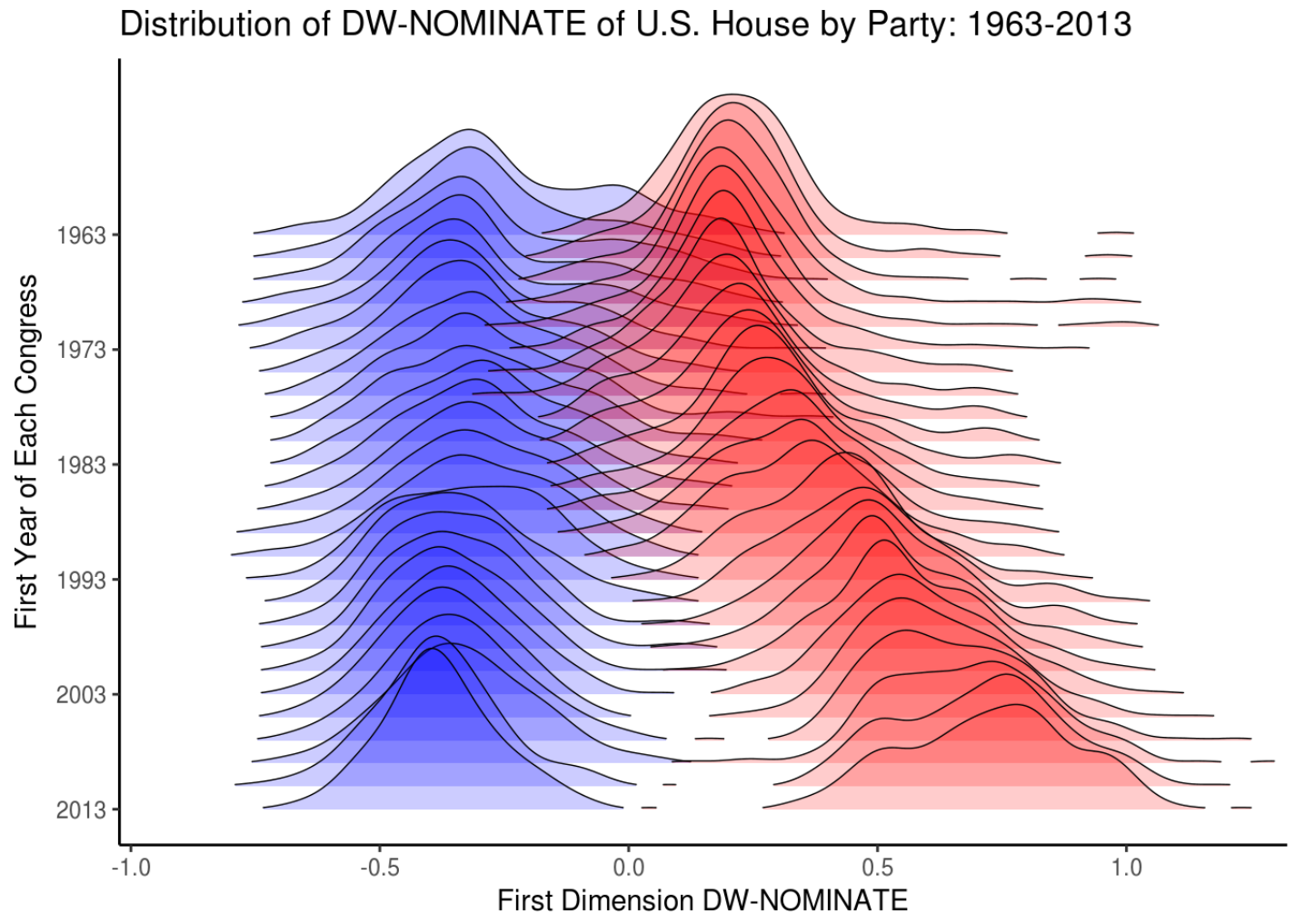


Multiple densities: Ridge plots

```
library(ggridges)

ggplot(filter(gapminder_2002,
              continent != "Oceania"),
       aes(x = lifeExp,
           fill = continent,
           y = continent)) +
  geom_density_ridges()
```

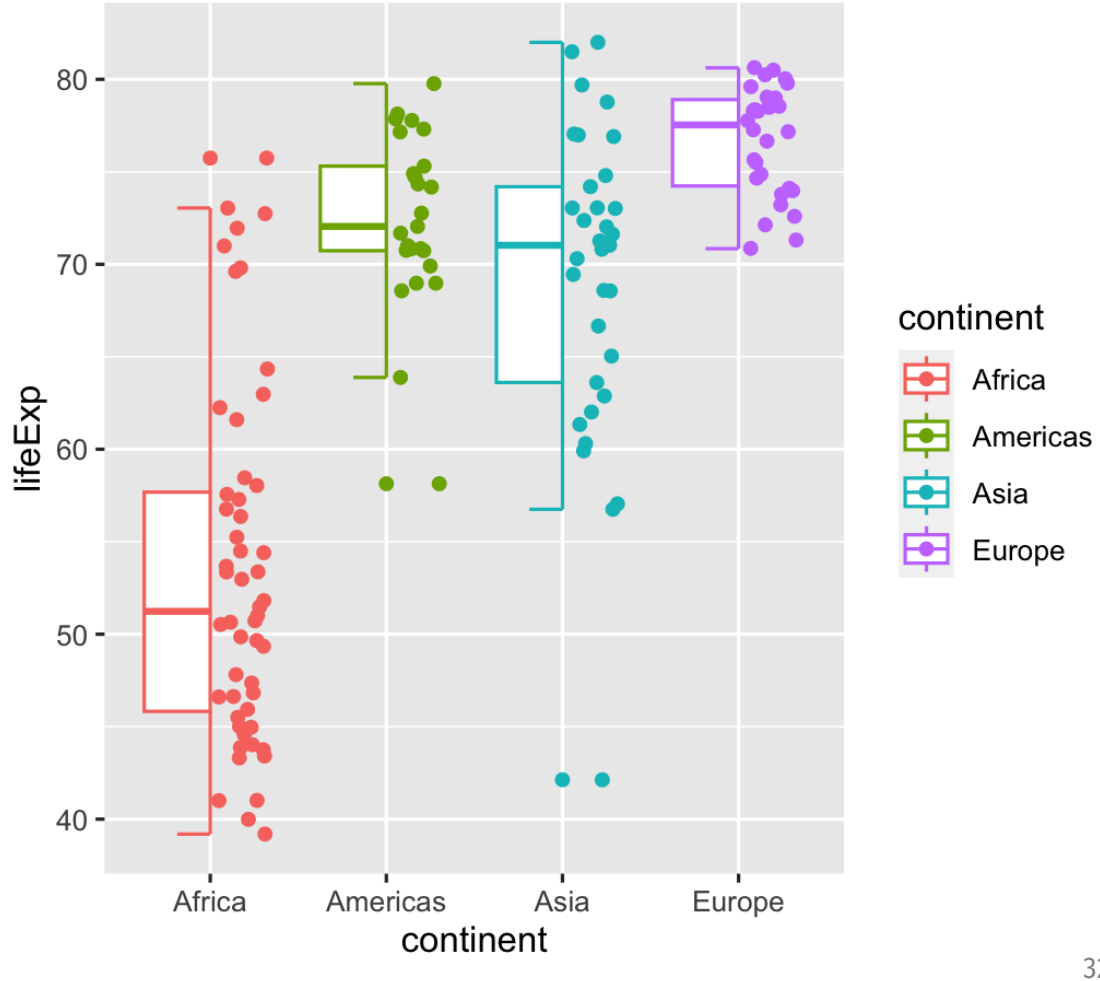


Multiple densities: Ridge plots



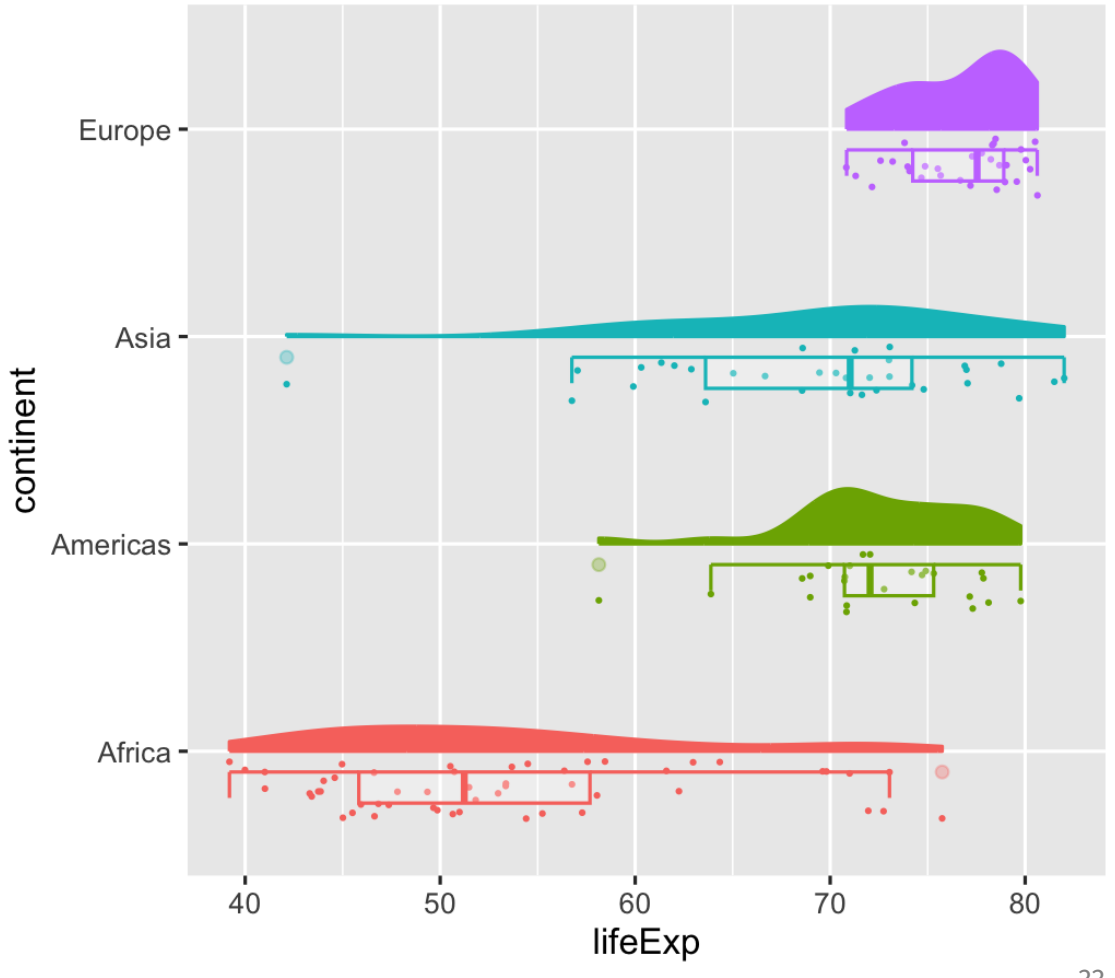
Multiple geoms: `gghalves`

```
library(gghalves)

ggplot(filter(gapminder_2002,
              continent != "Oceania"),
       aes(y = lifeExp,
           x = continent,
           color = continent)) +
  geom_half_boxplot(side = "l") +
  geom_half_point(side = "r")
```



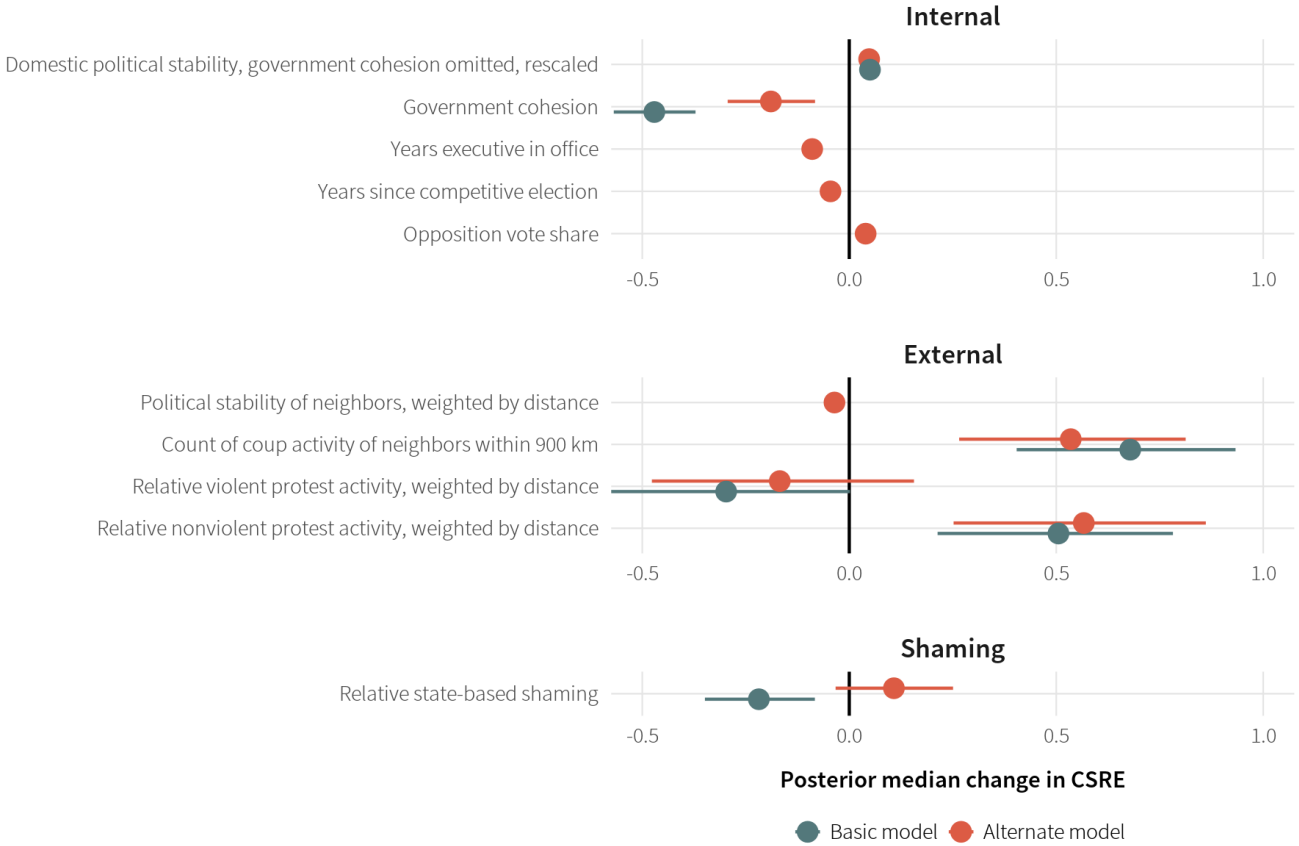
Multiple geoms: Raincloud plots

```
library(gghalves)

ggplot(filter(gapminder_2002,
              continent != "Oceania"),
       aes(y = lifeExp,
           x = continent,
           color = continent)) +
  geom_half_point(side = "l", size = 0.3) +
  geom_half_boxplot(side = "l", width = 0.5,
                    alpha = 0.3, nudge = 0.1)
  geom_half_violin(aes(fill = continent),
                   side = "r") +
  guides(fill = "none", color = "none") +
  coord_flip()
```

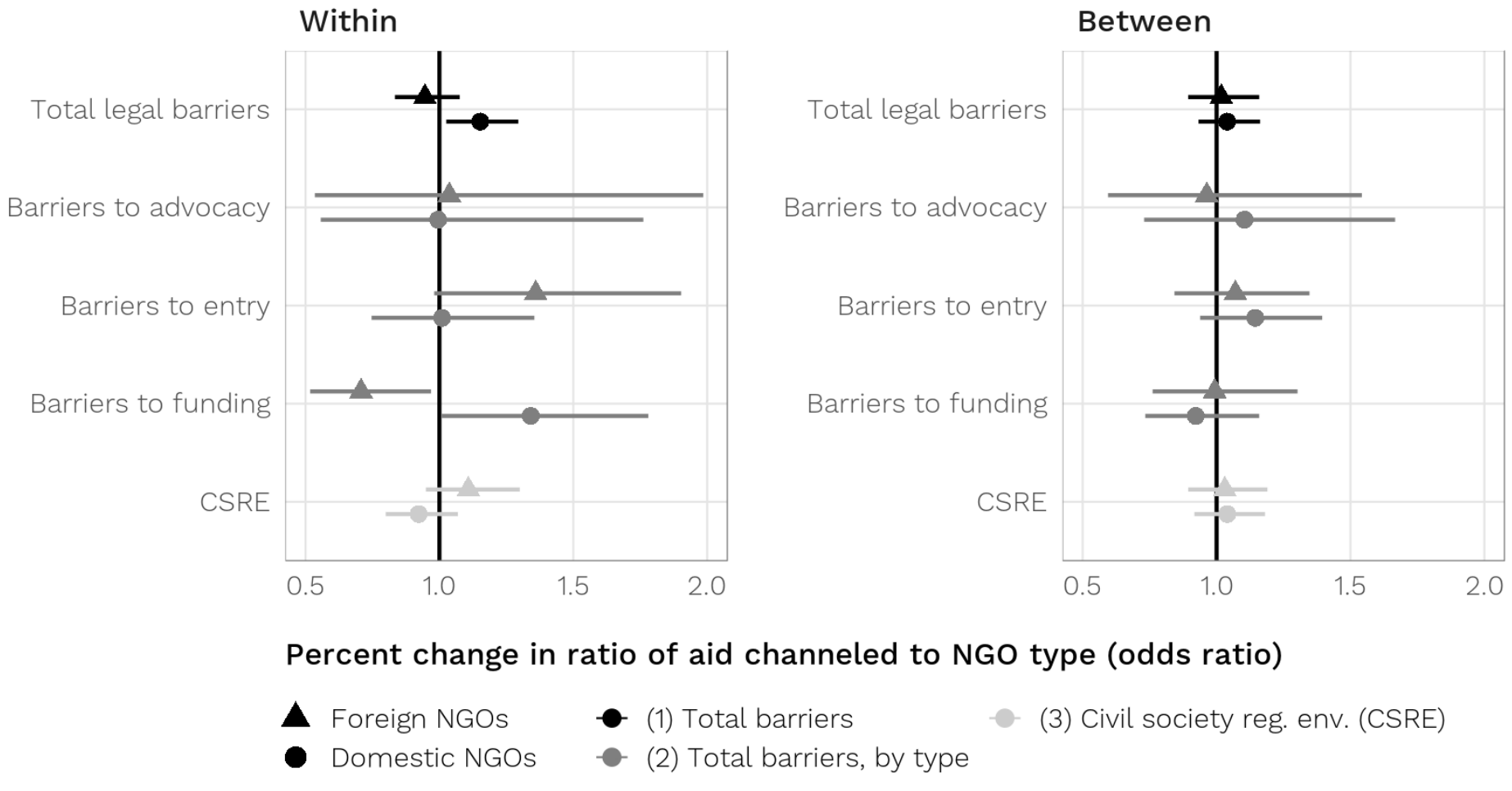


Uncertainty in model estimates

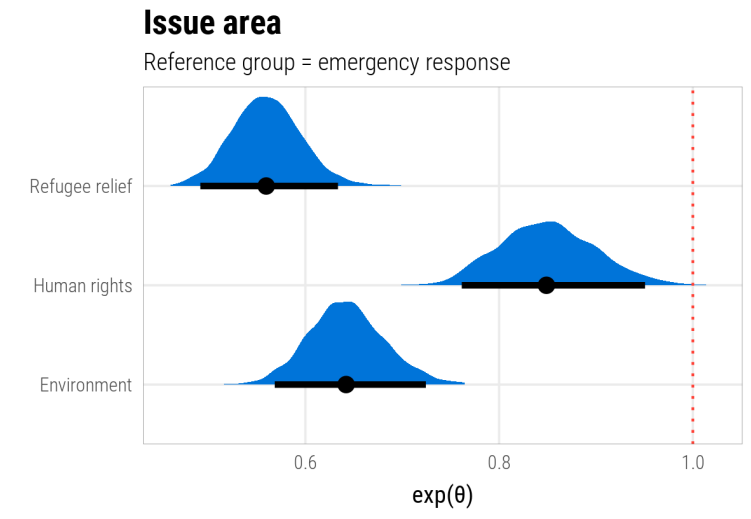
(You'll learn how to make these in the next session)

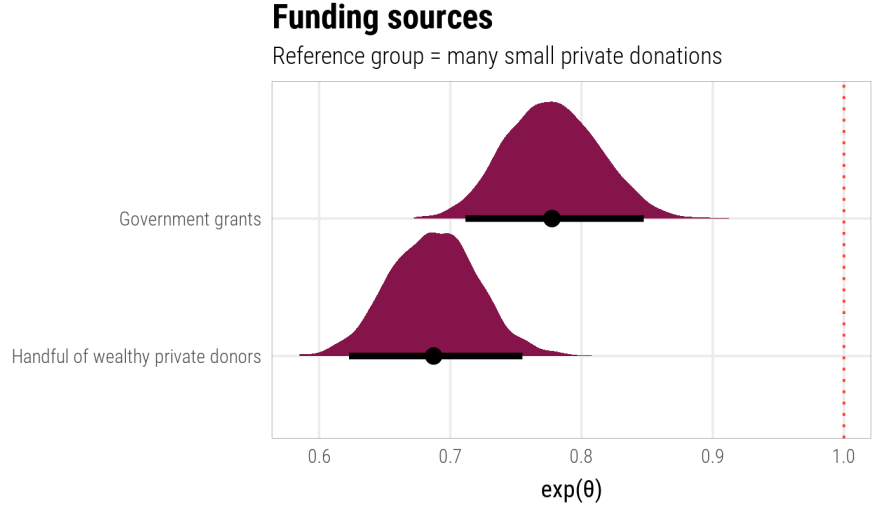


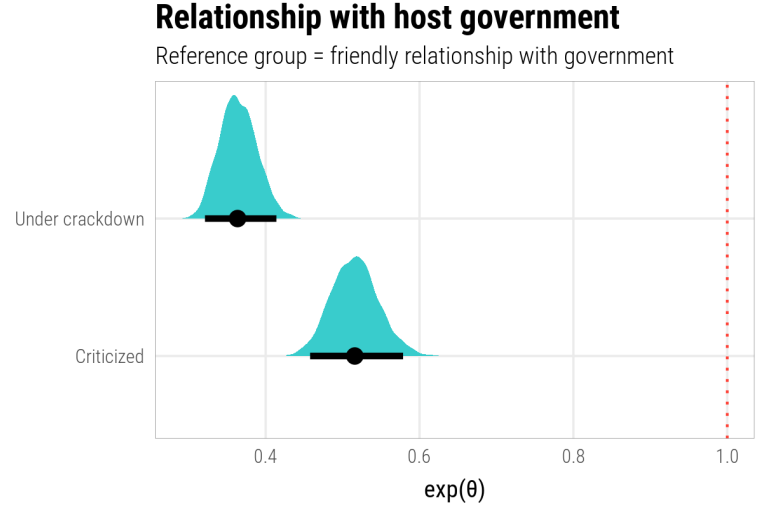
Uncertainty in model estimates



Uncertainty in model estimates

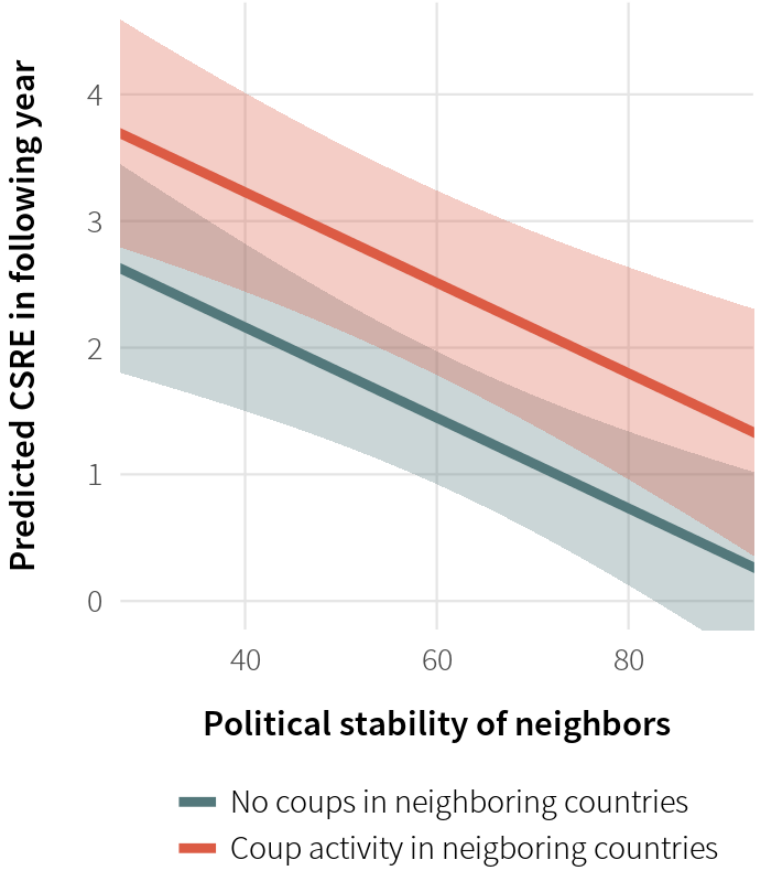


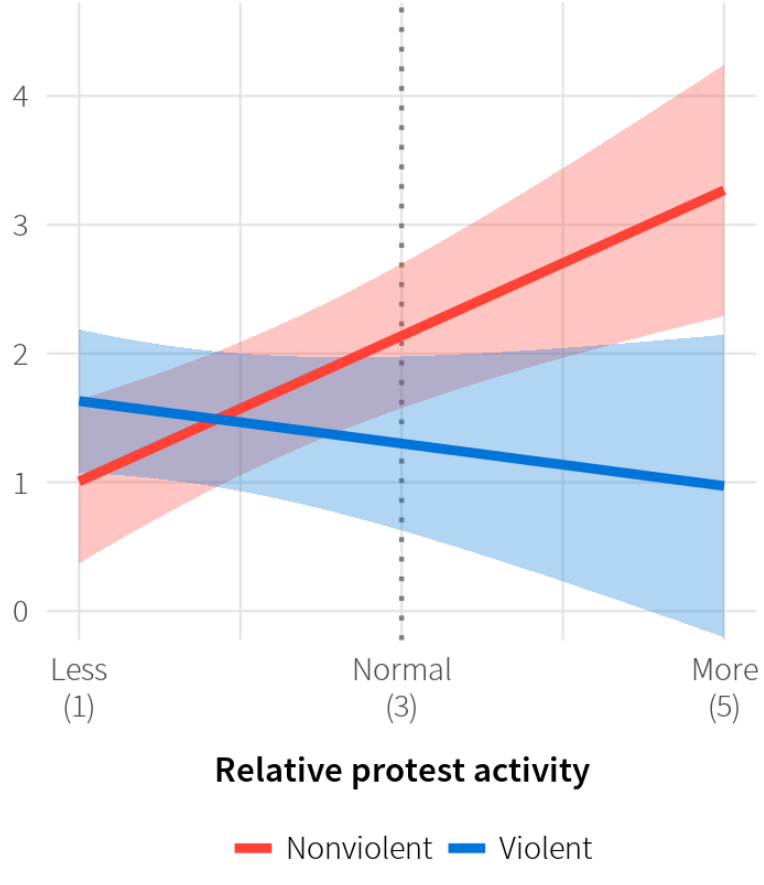




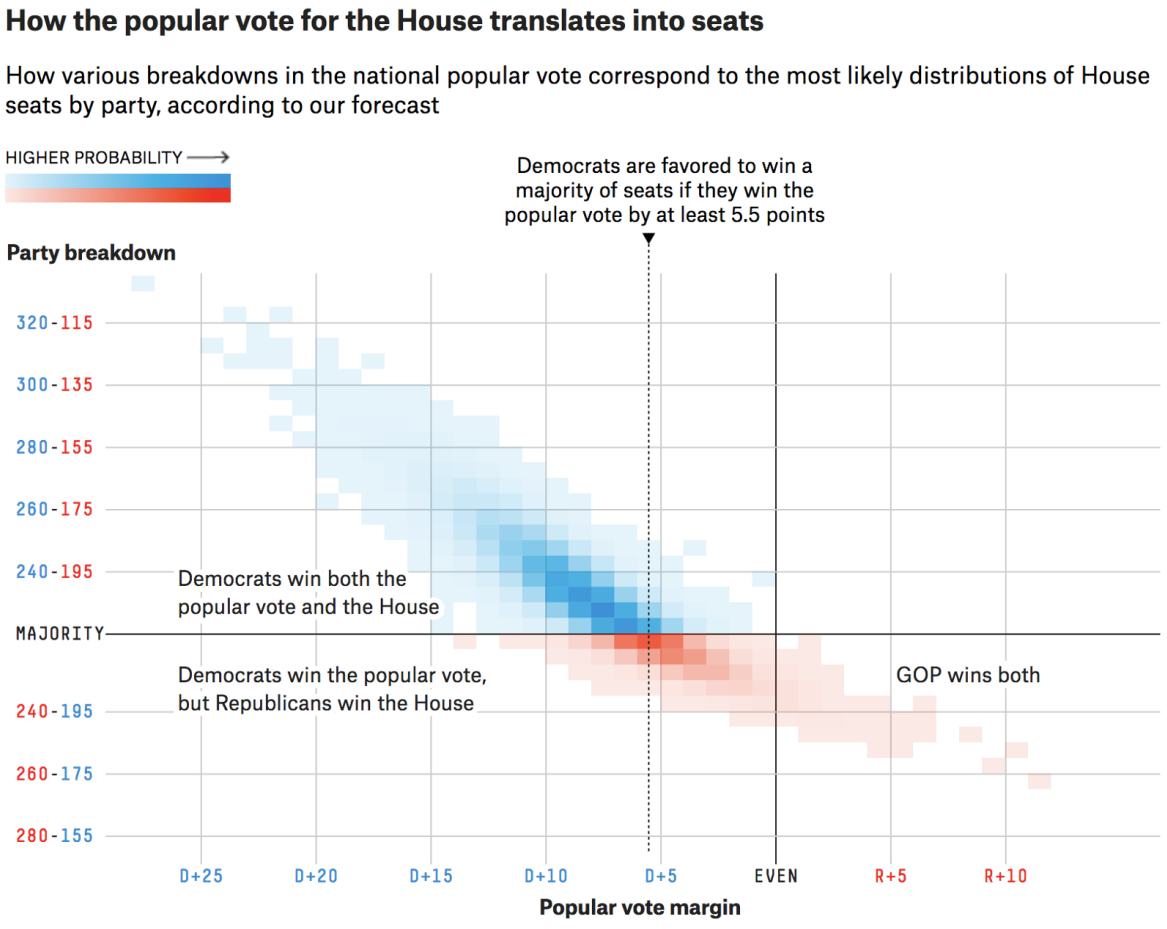
Uncertainty in model effects

(You'll learn how to make these in the next session)





Uncertainty in model outcomes



FiveThirtyEight's 2018 midterms model outcomes plot