

Session 6

PMAP 8921: Data Visualization with R Andrew Young School of Policy Studies Summer 2022

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!

1 in 5 vs. 20%







<1 in 100

Chance the Democrat wins (<0.1%)

>99 in 100

Chance the Republican wins (>99.9%)

Candidate	Forecasted vote share	Chance of winning
Mitt Romney (R)	59.6	>99 in 100 (>99.9%)
Jenny Wilson (D)	28.4	<1 in 100 (<0.1%)
Other candidates	12.0	<1 in 100 (<0.1%)
	10% 20 30 40 50 60 70	80 90 100





Chance the Democrat wins (25.5%)

3 in 4

Chance the Republican wins (74.5%)

Candidate	Forecasted vote share	Chance of winning
Ted Cruz (R) Incumbent	51.4	3 in 4 (74.5%)
Beto O'Rourke (D)	47.1	1 in 4 (25.5%)
Neal M. Dikeman (Lib.)	1.5	<1 in 100 (<0.1%)
	10% 20 30 40 50 60 70	80 90 100

Chance of rain = Probability × Area





SEVERITY Category 5 4 3 2 1 Tropical storm



Hurricane Maria map, New York Times

Hurricane Maria map, NOAA

The needle



The needle

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)
```



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



Density plots: Kernels and bandwidths

Different options for calculus change the plot shape



Density plots: Kernels and bandwidths

Different options for calculus change the plot shape



"epanechnikov"



"rectangular"





Show specific distributional numbers







Violin plots

Mirror density plot and flip

Often helpful to overlay other things on it



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



Multiple histograms

Facet with a different variable



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



Multiple densities: Ridge plots

library(ggridges)



Multiple densities: Ridge plots

Distribution of DW-NOMINATE of U.S. House by Party: 1963-2013



Multiple geoms: gghalves

library(gghalves)



Multiple geoms: Raincloud plots

continent

```
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



Percent change in ratio of aid channeled to NGO type (odds ratio)

Foreign NGOs
 (1) Total barriers
 (3) Civil society reg. env. (CSRE)
 (2) Total barriers, by type

Uncertainty in model estimates



Uncertainty in model effects

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

Uncertainty in model outcomes

How the popular vote for the House translates into seats

How various breakdowns in the national popular vote correspond to the most likely distributions of House seats by party, according to our forecast



FiveThirtyEight's 2018 midterms model outcomes plot