Data Visualization with ggplot

Election Data Science

Peter Licari

2020-09-28

What we'll be learning:

- The 3 main chart types (5 `geom_`s)
  1. Bar charts (`geom_bar` and `geom_hist`)
  2. Dot charts (`geom_dot`)
  3. Line charts (`geom_line` and `geom_smooth`)
- Different aesthetic options
  1. Color/Fill
  2. Size
  3. Transparency
  4. Positioning/Grouping
- Some helpful theme stuff

Let's grab the data:

```
library(tidyverse)
library(anesr)

# 2016 American National Election Survey
data(timeseries_2016)
anes16 <- timeseries_2016
rm(timeseries_2016)
```

First, some semantics.

```
ggplot(data, aes(x=x, y=y, color = z)) + geom_point()
ggplot() + geom_point(data, aes(x=x, y=y, color = z))
ggplot(data) + geom_point(aes(x=x, y=y, color = z))
```

These will all make the same images!

Bar Charts (geom_bar)

The Code | Output

```
to_na <- function(x){ifelse(x<0,NA,x)}



anes16a <- anes16 %>%
  select(V161003, V162034) %>%
  mutate(across(everything(), to_na)) %>%
  mutate(V162034 =  as.factor(V162034),
         V161003a = case_when(
    V161003 == 1 ~ "Always",
    V161003 == 2 ~ "Most of the time",
    V161003 == 3 ~ "About half the time",
    V161003 == 4 ~ "Some of the time",
    V161003 == 5 ~ "Never"
  ))

# plot
bar <- ggplot(anes16a, aes(x = V161003a)) + geom_bar()
```

Histogram (`geom_histogram`)

Code | Output

```
anes16b <- anes16 %>%
  mutate(V161267 = ifelse(V161267<0,NA,V161267))

hist <- ggplot(anes16b, aes(x = V161267)) + geom_histogram()
```

Dot chart (`geom_dot`)

Code Output

```
# Age, FT of Trump, 3pt Party ID Interest
anes16c <- anes16 %>%
  select(V161092, V161267, V161155, V161003) %>%
  mutate(across(everything(), to_na))

dot <- ggplot(data = anes16c, aes(x = V161267, y = V161092 )) + geom_po
```

Line Chart (geom_smooth)

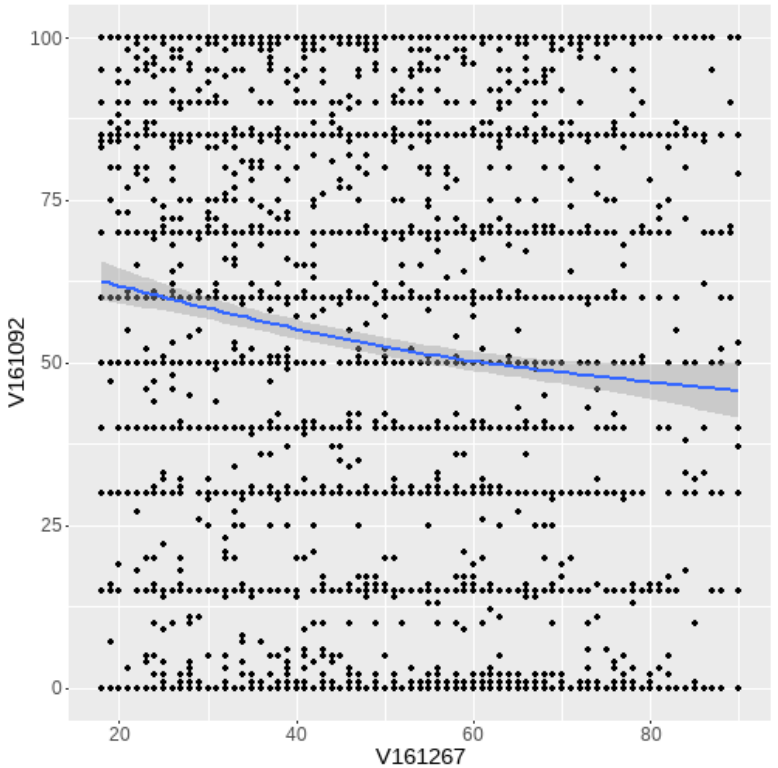
Code Output

```
anes16ca <- anes16 %>%
  select(V161092, V161267) %>%
  mutate(across(everything(), to_na))

smooth <- ggplot(data = anes16ca, aes(x = V161267, y = V161092 )) + geor
```

Bonus! Point + Smooth

```
dot + geom_smooth()
```




```
# ggplot(data = anes16c, aes(x = V161267, y = V161092 )) + geom_dot + ge
```

Line Chart (`geom_line`)

Code | Output

```
anes16d <- anes16 %>%
  select(V161267, V161086, V162079) %>%
  rename(pre = V161086, post = V162079) %>%
  mutate(across(everything(),to_na)) %>%
  summarise(`Pre-Election` = mean(pre, na.rm = T),
            `Post-Election` = mean(post, na.rm = T)) %>%
  pivot_longer(names_to = "xvar", cols = everything())

line <- ggplot(anes16d, aes(x = xvar, y = value)) + geom_line(group = "
```

Color aesthetic (`color`)

Code | Output

```
anes16c <- anes16c %>%
  filter(!(V161155 == 5), !(V161155 == 0)) %>%
  mutate(V161155 = as.factor(V161155))

dot1 <- ggplot(data = anes16c, aes(x = V161267, y = V161092, color = V16
```

Dot size

Code Output

```
anes16c <- anes16c %>%
  filter(!(V161155 == 5), !(V161155 == 0)) %>%
  mutate(V161155 = as.factor(V161155))

dot2 <- ggplot(data = anes16c, aes(x = V161267, y = V161092, size = V161
```

Transparency

Code Output

```
anes16c <- anes16c %>%
  filter(!(V161155 == 5), !(V161155 == 0)) %>%
  mutate(V161155 = as.factor(V161155))

dot3 <- ggplot(data = anes16c, aes(x = V161267, y = V161092, alpha = V16
```

Grouping

Code Output

```
anes16a <- anes16 %>%
  select(V161003, V162034) %>%
  mutate(across(everything(), to_na)) %>%
  mutate(V162034 =  as.factor(V162034),
         V161003a = case_when(
    V161003 == 1 ~ "Always",
    V161003 == 2 ~ "Most of the time",
    V161003 == 3 ~ "About half the time",
    V161003 == 4 ~ "Some of the time",
    V161003 == 5 ~ "Never"
  ))

# plot
bar1 <- ggplot(anes16a, aes(x = V161003a, fill=V162034)) + geom_bar()
```

Helpful theme stuff:

- `theme_`
- `theme`
- `coord_flip`
- `scale_MAP_TYPE`
- `labs`