Problem set 1

Solutions

Due 10/4/2021 at 5pm

Below I will include the most basic straightforward plots for all questions. Then I will include versions that have been improved in various ways.

Question 1: US presidential election results

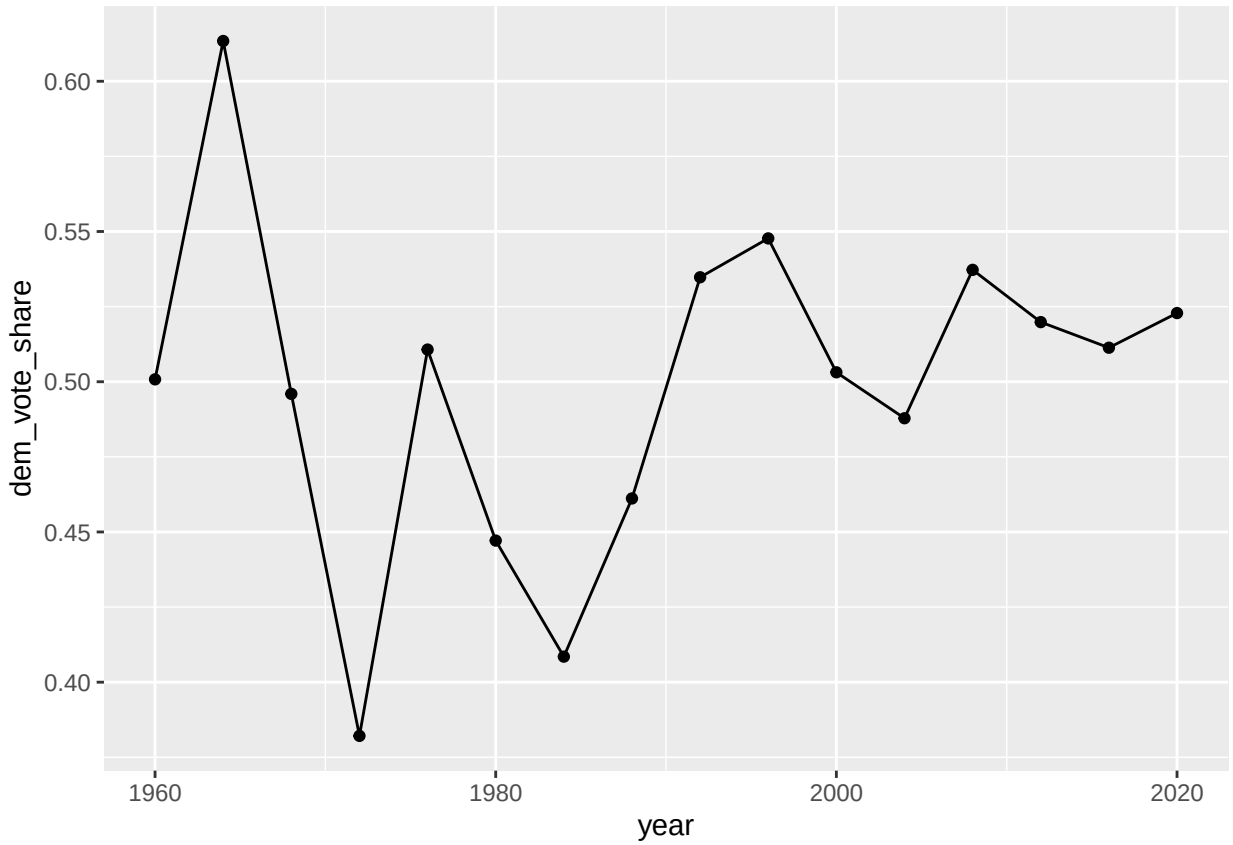
Load the data:

```
data_path <- "https://raw.githubusercontent.com/UChicago-pol-methods/IntroQSS-F21/main/data/"
df <- read_csv(str_c(data_path, "yearly_county_pres_results_wide.csv"))
```

This data includes `dem_vote_share` (the share of the two-party vote won by the Democratic candidate) and `dem_county_share` (the share of counties won by the Democratic candidate) in each U.S. presidential election since 1960.

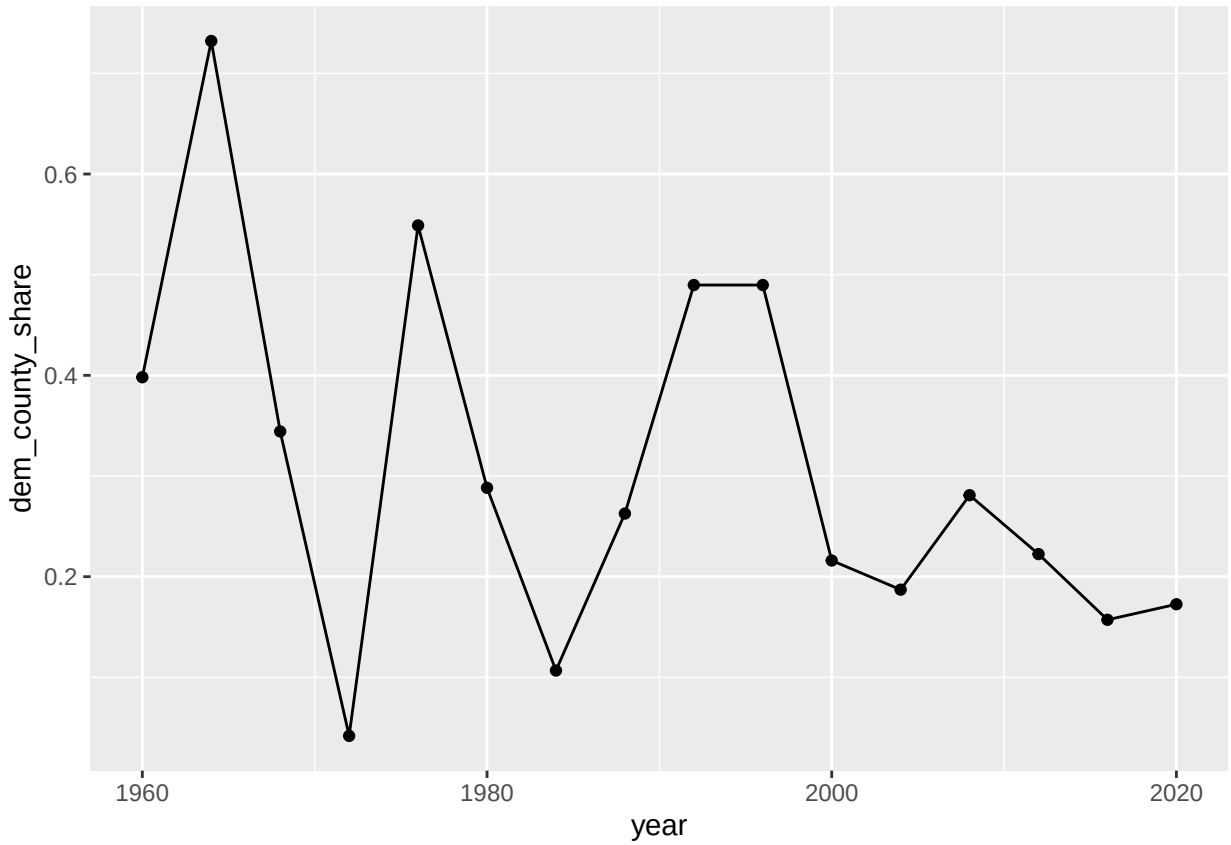
1a) Make a plot showing the Democratic vote share (vertical axis) in each year (horizontal axis). Draw a point for each year and connect them with a line.

```
df %>%
  ggplot(aes(x = year, y = dem_vote_share)) +
  geom_point() +
  geom_line()
```



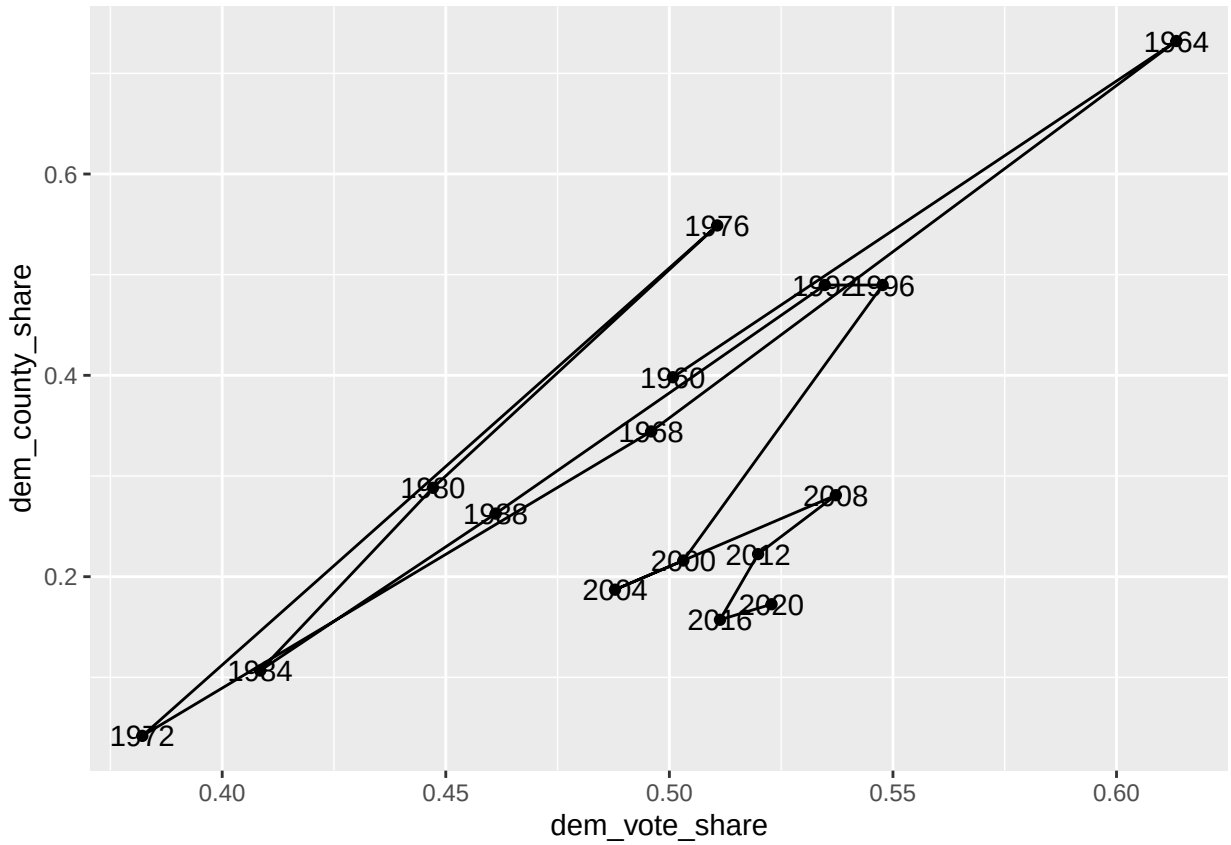
1b) Make a plot showing the Democratic county share (vertical axis) in each year (horizontal axis).

```
df %>%
  ggplot(aes(x = year, y = dem_county_share)) +
  geom_point() +
  geom_line()
```



1c) Now make a plot showing Democratic county share (vertical axis) and Democratic vote share (horizontal axis), again connecting the points with a line. (Hint: use `geom_path()`.) Label each point with the corresponding year. (Hint: use `geom_text()`.)

```
df %>%
  ggplot(aes(x = dem_vote_share, y = dem_county_share, label = year)) +
  geom_point() +
  geom_path() +
  geom_text()
```

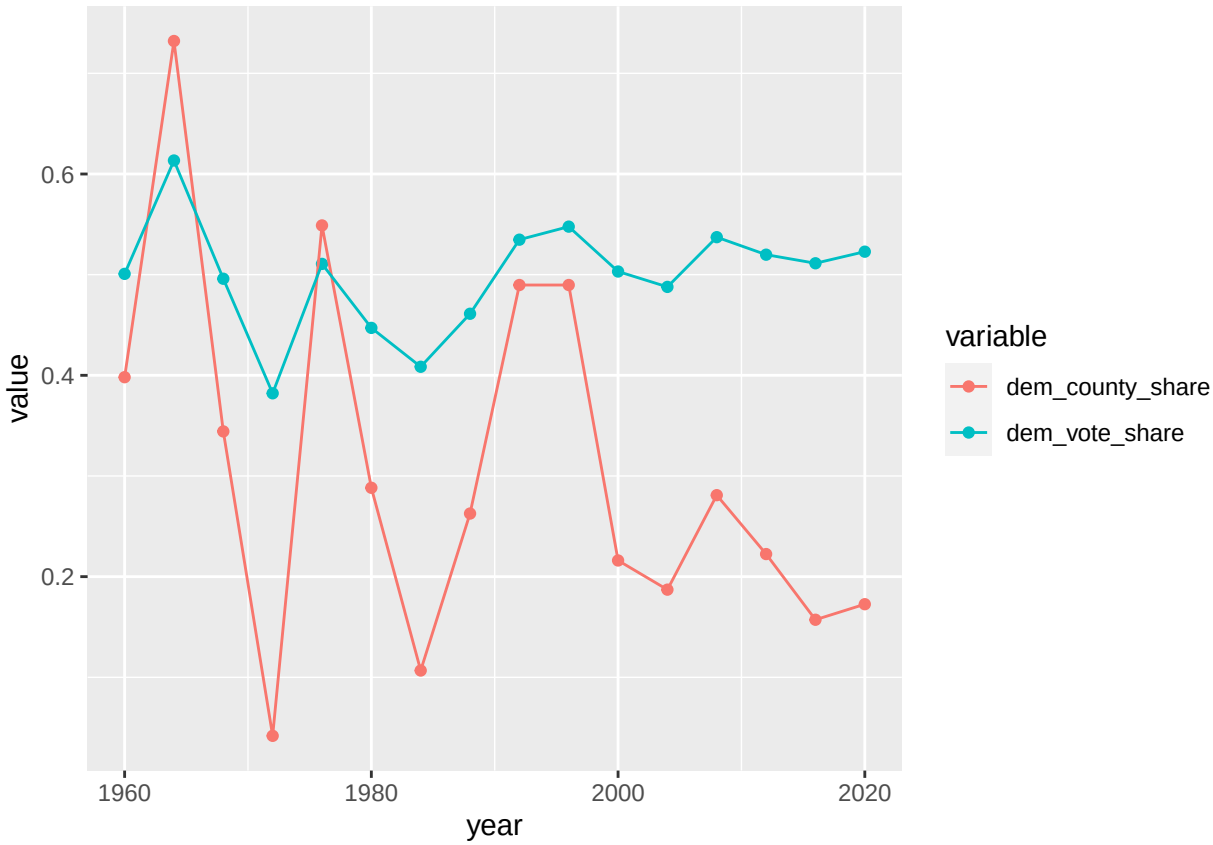


Now load this dataset, which is the same data organized differently:

```
df2 <- read_csv(str_c(data_path, "yearly_county_pres_results_long.csv"))
```

1d) Make a plot showing both the Democratic vote share and Democratic county share (vertical axis) in each year (horizontal axis), with a different color for each series.

```
ggplot(df2, aes(x = year, y = value, col = variable)) +
  geom_point() +
  geom_line()
```



1e) Do you prefer the plot that shows county share (y) against vote share (x), or the one that shows both county share and vote share (y) against time (x)? Why?

The two plots contain exactly the same information, so preferences between them are all about the viewer's perception and expectations. Many people expect time to be along the x-axis, so the second plot may be easier to understand. The first plot emphasizes that the county share and vote share used to covary more than they have in recent elections. The second plot emphasizes the passage of time, and makes it easier to see that a discrepancy has emerged between the share of votes vs the share of counties won by Democrats. The reason for this discrepancy, incidentally, is that Democratic support has increasingly become concentrated in cities (and therefore in the counties where cities are located).

Question 2: democracy and GDP

Load the data. This is an extract from the V-Dem dataset (https://www.v-dem.net/). Variables include `country_name` and `continent_name` (self-explanatory), `polyarchy` (V-Dem's measure of democracy), `pop` (World Bank measure of population), and `gdppc` (GDP per capita), all from 2010. The full dataset (available in the `vdemdata` `R` package) contains many more variables and years.

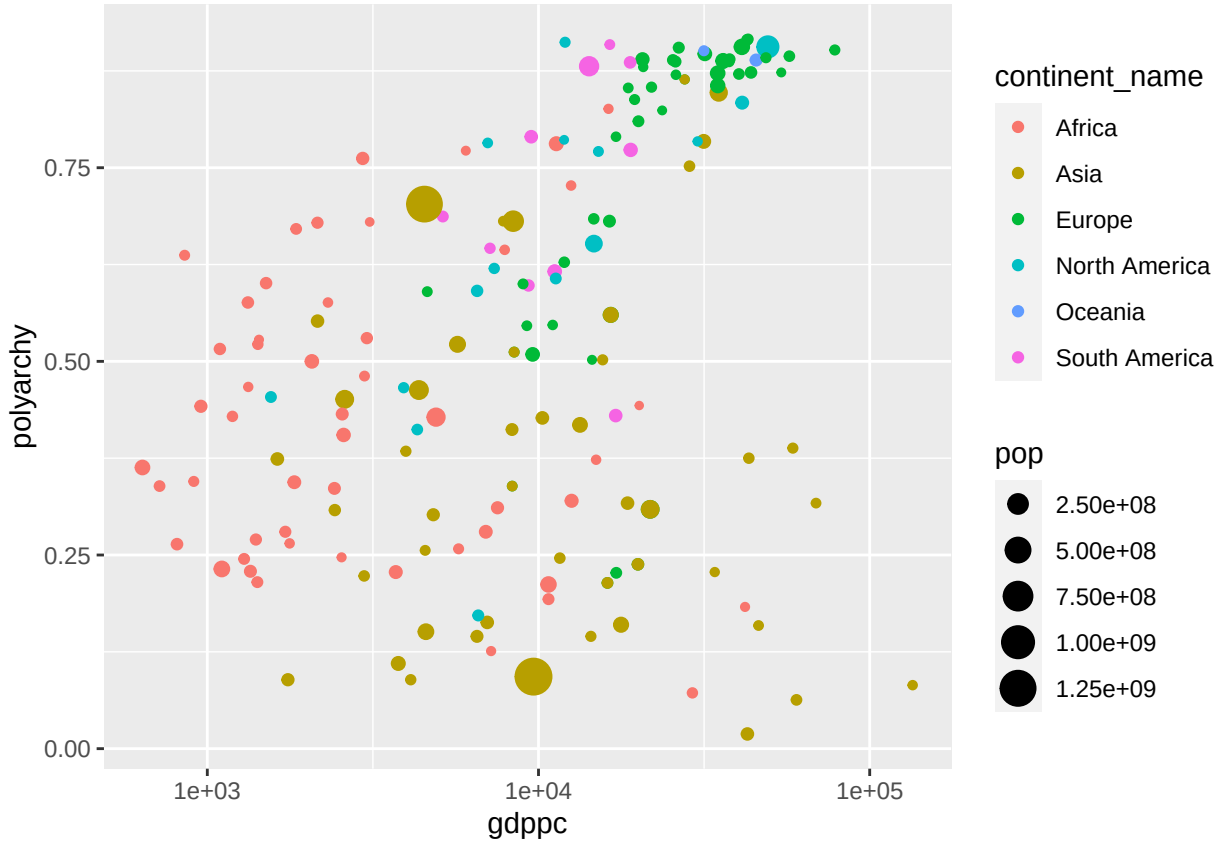
```
vd <- read_csv(str_c(data_path, "vdem_2010_extract.csv"))
```

2a) Make a scatterplot of the V-Dem polyarchy score (vertical axis) against GDP per capita (horizontal axis). Make the color of the dots reflect the continent, and the size reflect the population. Show the horizontal axis on the log scale.

```
p1 <- ggplot(vd, aes(x = gdppc, y = polyarchy, col = continent_name, size = pop)) +
  geom_point() +
```

```
  scale_x_log10()
p1
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```

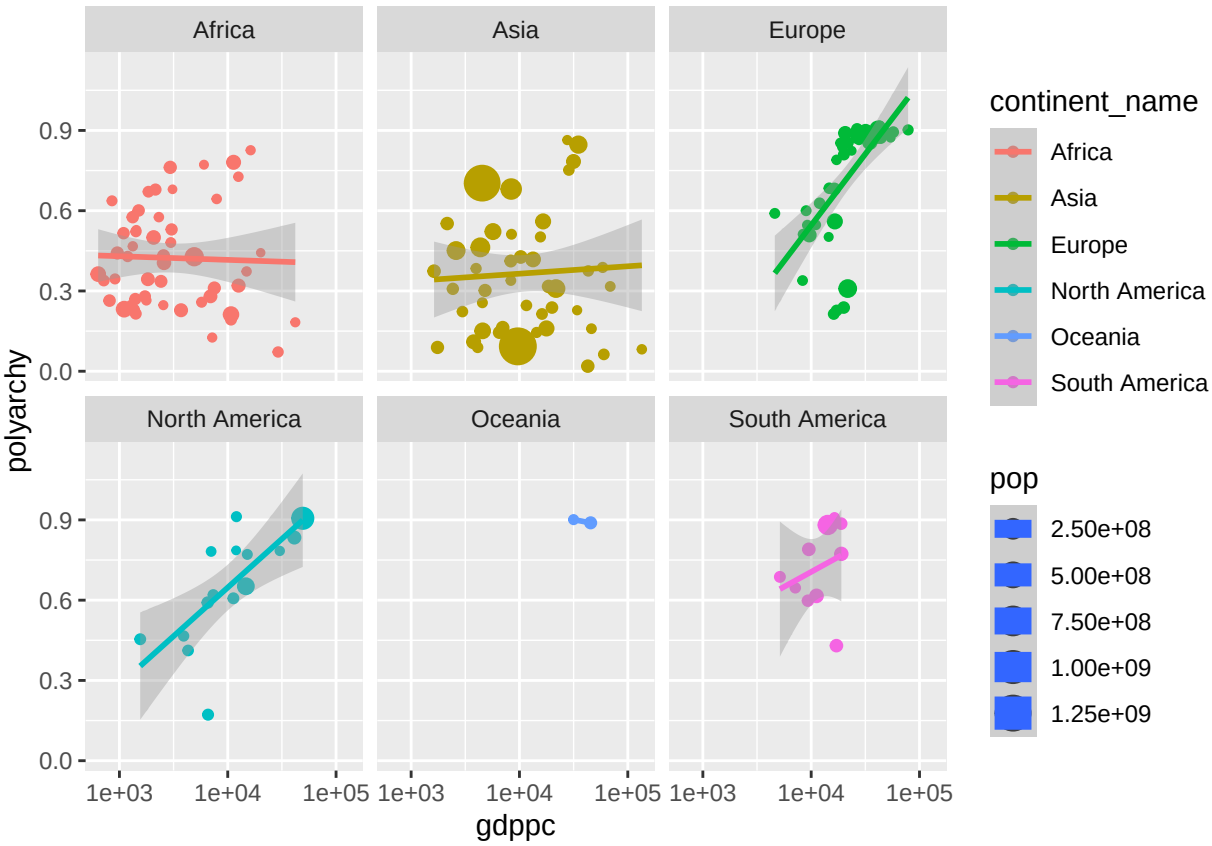


2b) Now make the same figure faceted by continent. Add a linear regression line (use `geom_smooth(method = lm)`). How does the relationship between GDP per capita and democracy differ across continents?

Again here's the most basic plot:

```
p1 +
  facet_wrap(vars(continent_name)) +
  geom_smooth(method = lm)
```

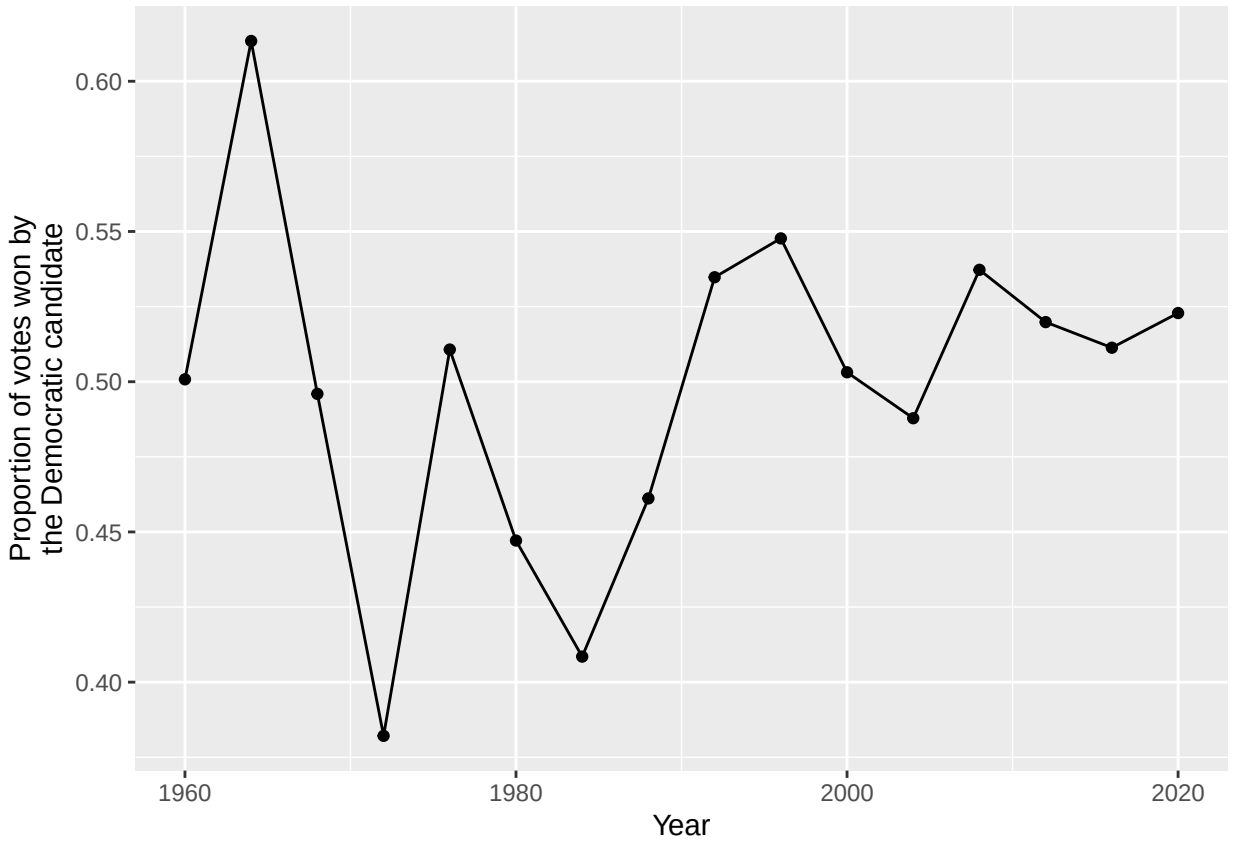
```
## `geom_smooth()` using formula 'y ~ x'
```



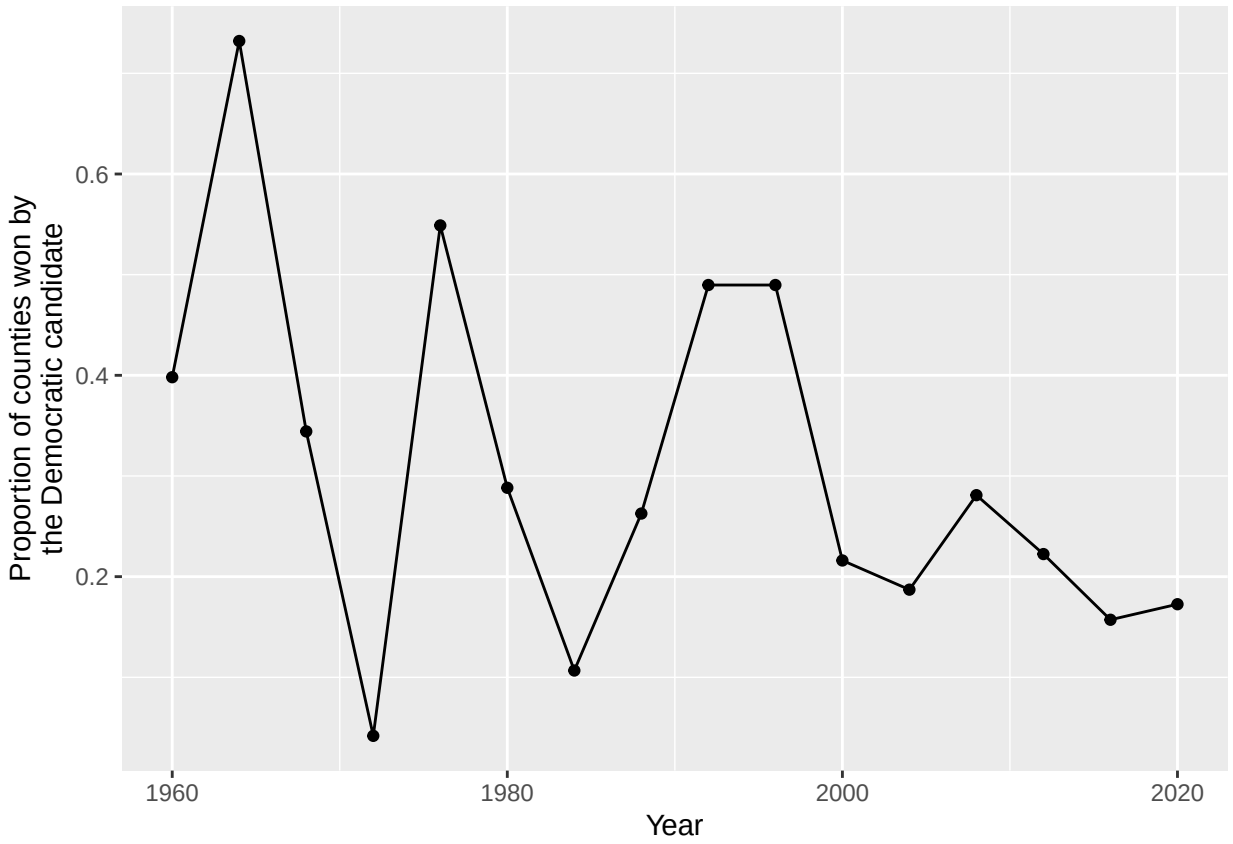
The interesting pattern is that within Europe and North America there is a positive relationship between wealth and democracy; within Africa and Asia there is no relationship.

Nicer plots

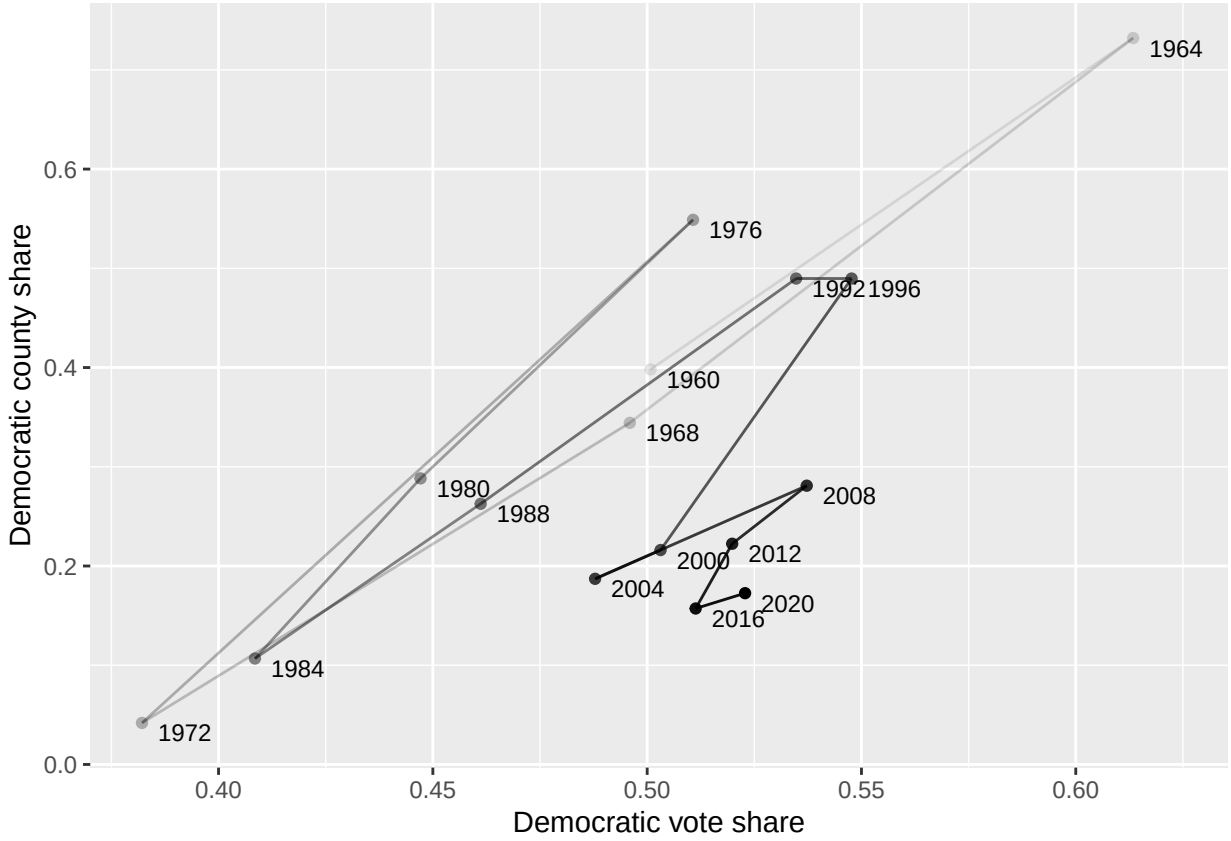
```
df %>%
  ggplot(aes(x = year, y = dem_vote_share)) +
  geom_point() +
  geom_line() +
  labs(x = "Year", y = "Proportion of votes won by\nthe Democratic candidate")
```



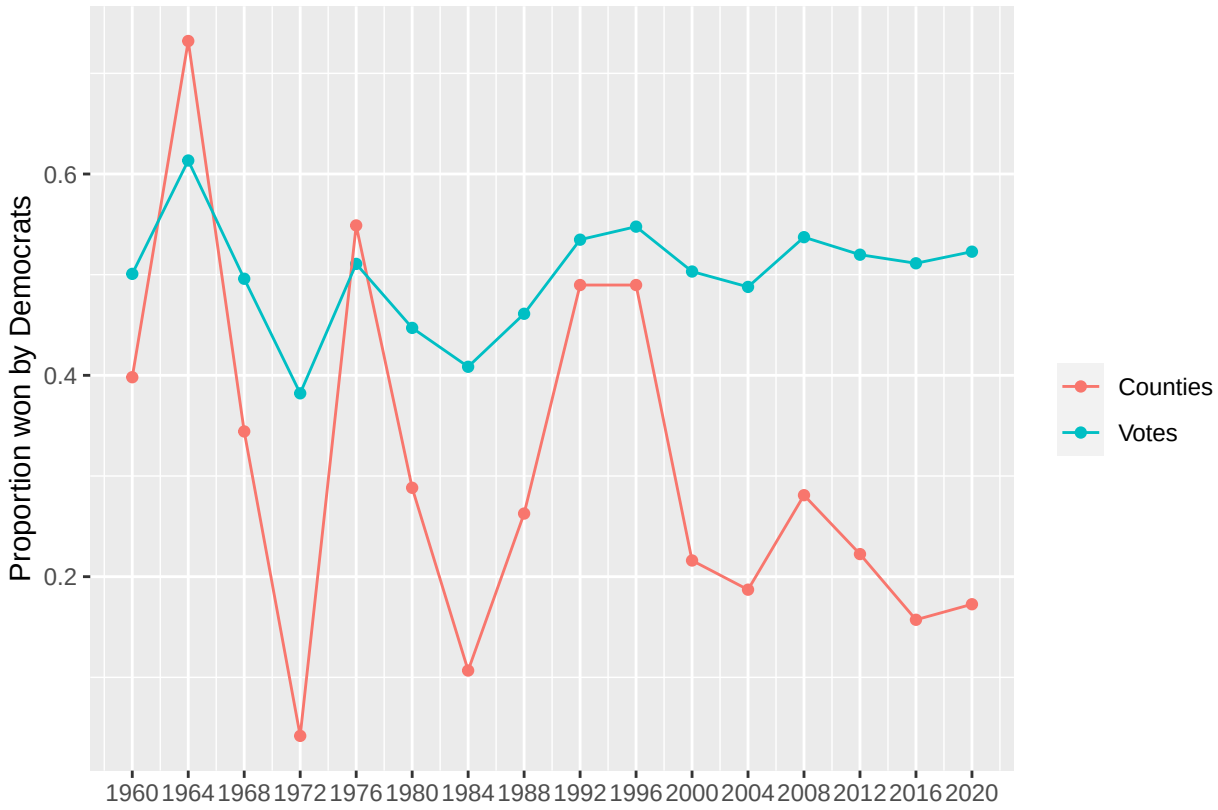

```
df %>%
  ggplot(aes(x = year, y = dem_county_share)) +
  geom_point() +
  geom_line() +
  labs(x = "Year", y = "Proportion of counties won by\nthe Democratic candidate")
```




```
df %>%
  ggplot(aes(x = dem_vote_share, y = dem_county_share, label = year)) +
  geom_point(aes(alpha = year), show.legend = F) +
  geom_path(aes(alpha = year), show.legend = F) +
  geom_text(nudge_x = .01, nudge_y = -.01, size = 3) +
  labs(x = "Democratic vote share", y = "Democratic county share")
```

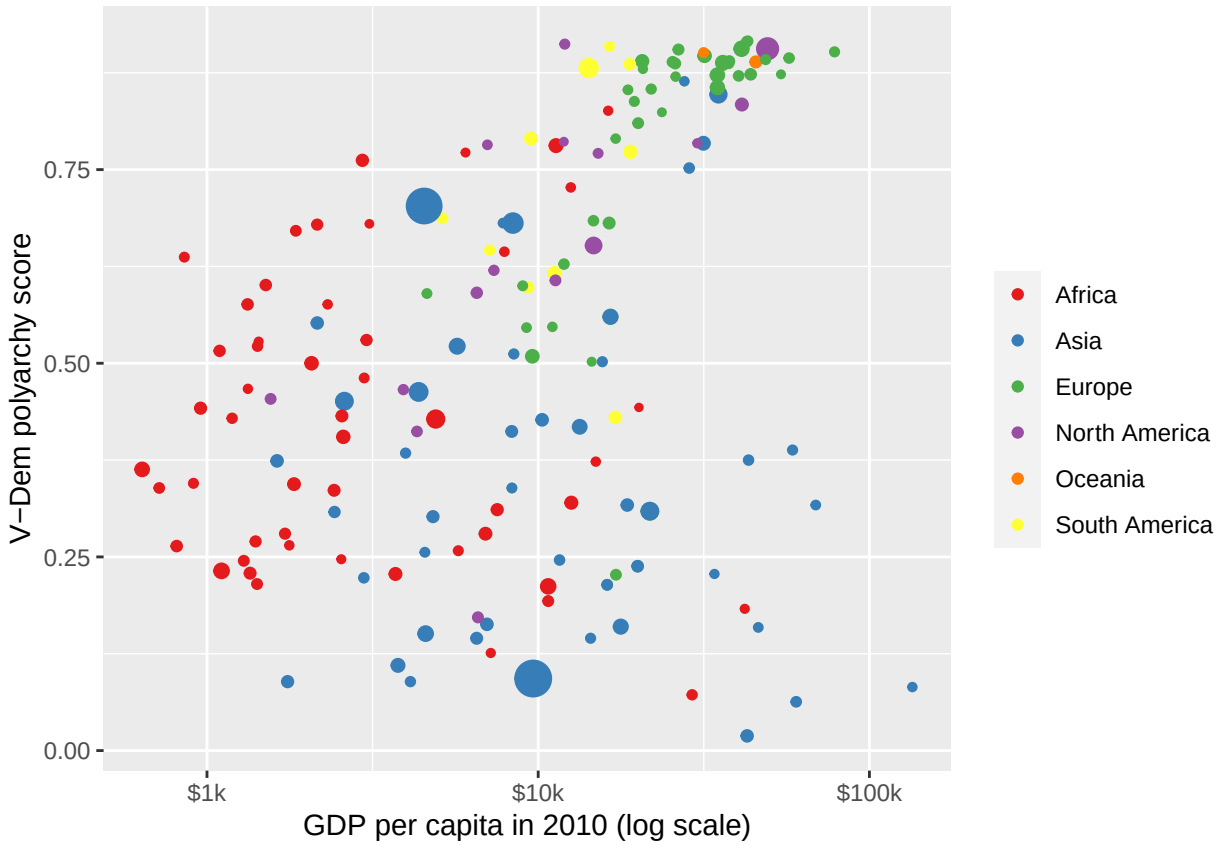



```
df2 %>%
  mutate(variable = ifelse(variable == "dem_county_share", "Counties", "Votes")) %>%
  ggplot(aes(x = year, y = value, col = variable)) +
  geom_point() +
  geom_line() +
  scale_x_continuous(breaks = seq(1960, 2020, by = 4)) +
  labs(x = "", y = "Proportion won by Democrats", col = "")
```



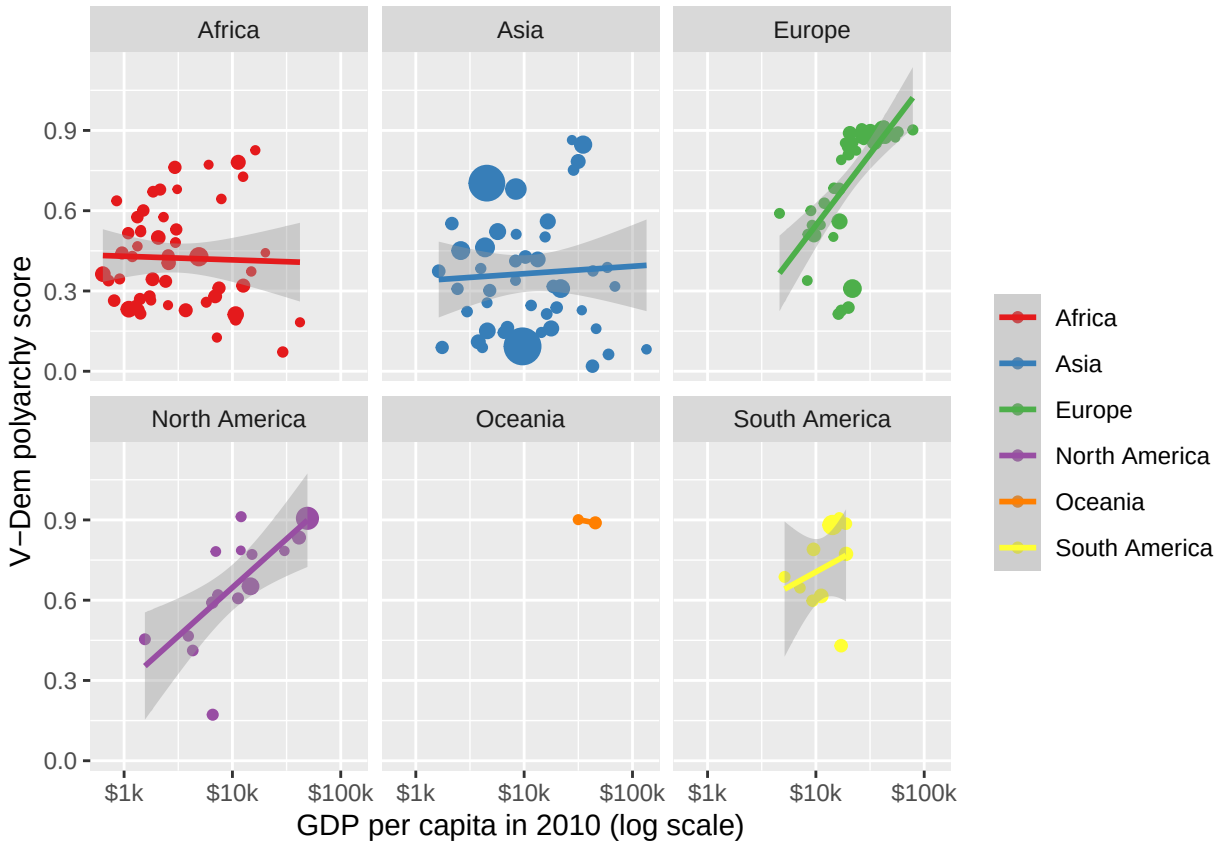

```
p2 <- ggplot(vd, aes(x = gdppc, y = polyarchy, col = continent_name, size = pop)) +
  geom_point() +
  scale_x_log10(breaks = c(1e+3, 1e+4, 1e+5), labels = c("$1k", "$10k", "$100k")) +
  scale_color_brewer(palette = "Set1") +
  scale_size_continuous(guide = "none") +
  labs(x = "GDP per capita in 2010 (log scale)", y = "V-Dem polyarchy score", col = "")
p2
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```




```
p2 +
  facet_wrap(vars(continent_name)) +
  geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```


Africa
Asia
Europe
North America
Oceania
South America
V−Dem polyarchy score
0.0
0.3
0.6
0.9
$1k
$10k
$100k
GDP per capita in 2010 (log scale)
Africa
Asia
Europe
North America
Oceania
South America