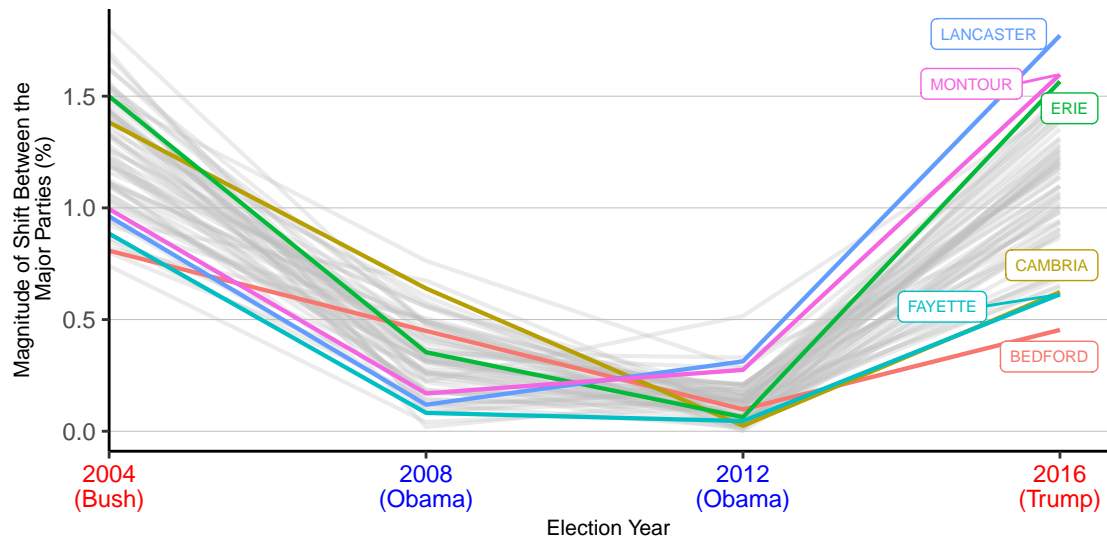


Breaking Down the Blue Wall in Pennsylvania

Visualizing the factors that led to a Republican victory in the state in the 2016 presidential election

Two-Party Voting Shifts in Pennsylvania Counties (2004–2016)

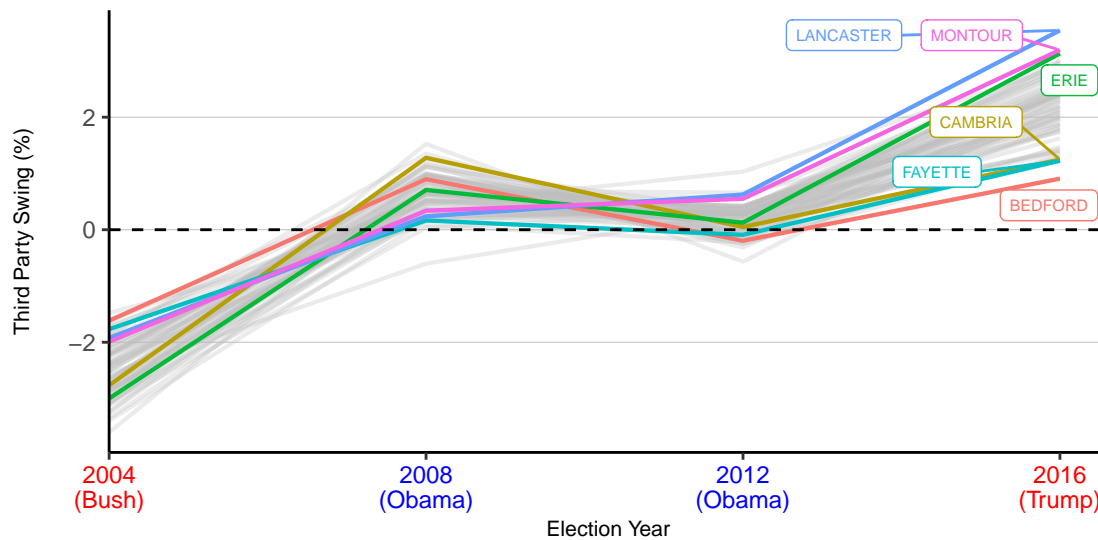
Every Pennsylvania county experienced major shifts in voting preferences from 2012 to 2016 compared to previous election cycles, regardless of party affiliation



Data source: Harvard Dataverse

Third Party Voting Shifts in Pennsylvania Counties (2004–2016)

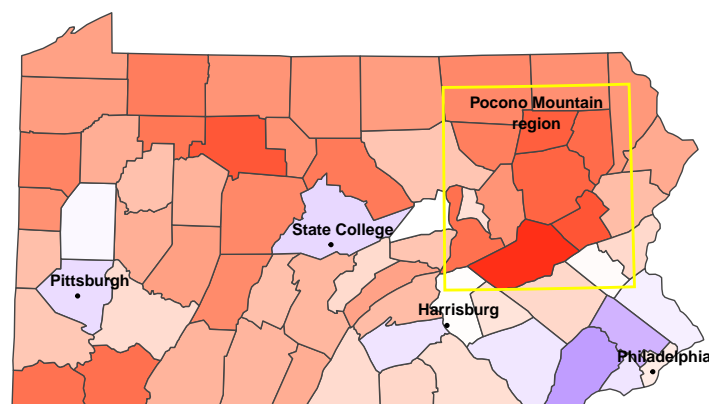
The same counties that experienced a large two-party shift in 2016 also experienced a comparatively large third party vote presence in 2016 than in previous election years.



Data source: Harvard Dataverse

Mapping Republican Vote Share Increase from 2012 to 2016

Republican vote share increased primarily in non-metropolitan counties and, on average, at a higher magnitude than did Democratic vote share anywhere. The Pocono Mountain region experienced the highest increase.



Increase in Republican Vote Share from 2012 to 2016 (%)

-15 -10 -5 0 5 10 15

Data source: Harvard Dataverse

Infographic Project Draft

Andrew Shih

Style Guide: tidyverse

2024-10-12

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
pa_election_df <- read.csv("~/Documents/MADS/Fa24/36-613 (Data Visualization)/Infographic Pro
  filter(state == "PENNSYLVANIA") |>
  filter(party == "DEMOCRAT" | party == "REPUBLICAN") |>
  group_by(year, state, county_name, office, candidate, party) |>
  summarize(
    new_candidatevotes = sum(candidatevotes),
    new_totalvotes = sum(totalvotes)
  ) |>
  select(
    year,
    state,
    county_name,
    party,
    new_candidatevotes,
    new_totalvotes
  ) |>
```

```
mutate(percentage_of_vote = new_candidatevotes * 100 / new_totalvotes) |>
arrange(state, county_name, party, year) |>
group_by(state, county_name, party) |>
mutate(previous_percentage = lag(percentage_of_vote)) |>
mutate(swing = percentage_of_vote - previous_percentage)
```

`summarise()` has grouped output by 'year', 'state', 'county_name', 'office', 'candidate'. You can override using the `.groups` argument.

Adding missing grouping variables: `office`, `candidate`

```
pa_election_df_swing <- pa_election_df |>
  select(
    year, state, county_name, party,
    percentage_of_vote, previous_percentage, swing
  ) |>
  group_by(county_name, year) |>
  summarize(two_party_swing = (sum(swing[party == "REPUBLICAN"], na.rm = TRUE) +
    sum(swing[party == "DEMOCRAT"], na.rm = TRUE)) / 2)
```

`summarise()` has grouped output by 'county_name'. You can override using the `.groups` argument.

```
pa_election_df_swing_avg <- pa_election_df_swing |>
  filter(year != 2000) |>
  group_by(county_name) |>
  summarize(avg = mean(abs(two_party_swing)))
```

```
library(gghighlight)
```

Loading required package: ggplot2

```
library(ggrepel)

county_list <- c(
  "LANCASTER",
  "MONTGOMERY",
  "ERIE",
  "CAMBRIA",
```

```

    "FAYETTE",
    "BEDFORD"
  )

plot_1 <- pa_election_df_swing |>
  filter(!year %in% c(2000, 2020)) |>
  filter(county_name != "CAMERON") |>
  ggplot() +
  geom_line(aes(x = year, y = abs(two_party_swing), color = county_name),
    size = 0.75, alpha = 1
  ) +
  gghighlight(
    county_name %in% county_list,
    label_params = list(size = 2),
    unhighlighted_params = list(size = 0.5, alpha = 0.3)
  ) +
  labs(
    title = "Two-Party Voting Shifts in Pennsylvania Counties (2004-2016)",
    subtitle = "Every Pennsylvania county experienced major shifts in voting
    preferences from 2012 to 2016 compared to previous election cycles,
    regardless of party affiliation",
    x = "Election Year",
    y = "Magnitude of Shift Between the
    Major Parties (%)",
    caption = "Data source: Harvard Dataverse"
  ) +
  scale_x_continuous(
    breaks = seq(2004, 2016, by = 4),
    labels = c(
      "2004\n(Bush)", "2008\n(Obama)",
      "2012\n(Obama)", "2016\n(Trump)"
    ),
    limits = c(2004, 2016),
    expand = expansion(mult = c(0, 0.05))
  ) +
  theme_classic() +
  theme(
    plot.title = element_text(
      margin = margin(b = 5),
      hjust = 0.5,
      face = "bold",
      size = 11.5
    )
  )

```

```

),
plot.subtitle = element_text(hjust = 0.5, size = 9.5),
plot.caption = element_text(size = 6),
plot.margin = margin(l = 120),
legend.position = "none",
panel.grid.major.y = element_line(color = "gray", size = 0.15),
panel.grid.minor.y = element_blank(),
panel.grid.major.x = element_blank(),
panel.grid.minor.x = element_blank(),
axis.text.x = element_text(color = c("red", "blue", "blue", "red")),
axis.title = element_text(size = 7.5)
)

```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
 i Please use `linewidth` instead.

Warning: Tried to calculate with group_by(), but the calculation failed.
 Falling back to ungrouped filter operation...

label_key: county_name

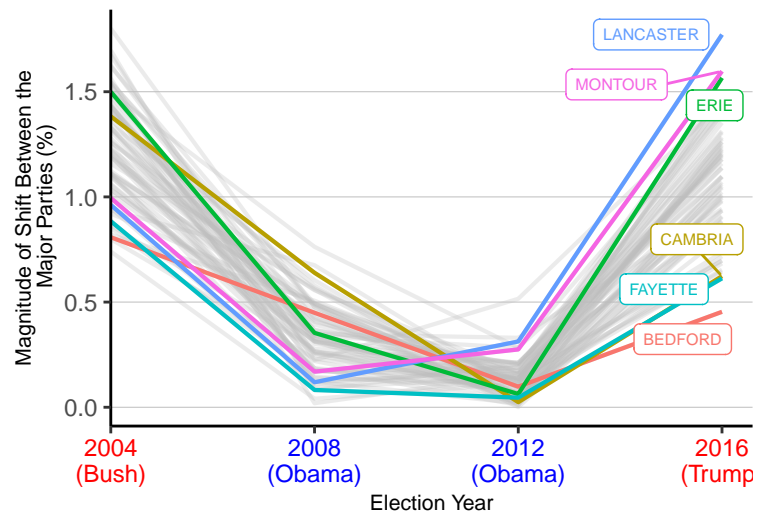
Warning: Vectorized input to `element_text()` is not officially supported.
 i Results may be unexpected or may change in future versions of ggplot2.

Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
 i Please use the `linewidth` argument instead.

plot_1

Two-Party Voting Shifts in Pennsylvania Counties (2)

Every Pennsylvania county experienced major shifts in voter preferences from 2012 to 2016 compared to previous election regardless of party affiliation



Data source: Harvard Dataverse

```
library(maps)
```

```
pa_election_df_rep <- pa_election_df |>
  filter(year %in% c(2012, 2016)) |>
  filter(party == "REPUBLICAN") |>
  select(-one_of(c("previous_percentage", "swing"))) |>
  group_by(county_name) |>
  summarize(inc_rep_vote_share = (sum(percentage_of_vote[year == 2016]) -
    sum(percentage_of_vote[year == 2012])))
```

```
pa_election_df_third <- read.csv("~/Documents/MADS/Fa24/36-613 (Data Visualization)/Infograph")
# filter(year %in% c(2012, 2016)) |>
filter(state == "PENNSYLVANIA") |>
filter(!party %in% c("DEMOCRAT", "REPUBLICAN")) |>
group_by(year, state, county_name, office, candidate, party) |>
summarize(
  new_candidatevotes = sum(candidatevotes),
  new_totalvotes = sum(totalvotes)
) |>
select(
  year,
  state,
  county_name,
```

```

    party,
    new_candidatevotes,
    new_totalvotes
  ) |>
  group_by(county_name, year) |>
  mutate(
    new_candidatevotes = sum(new_candidatevotes),
    new_totalvotes = first(new_totalvotes)
  ) |>
  filter(!(year == 2000 & candidate == "OTHER")) |>
  arrange(county_name, year, county_name) |>
  mutate(percentage_of_vote = new_candidatevotes * 100 / new_totalvotes) |>
  group_by(state, county_name) |>
  mutate(previous_percentage = lag(percentage_of_vote)) |>
  mutate(swing = percentage_of_vote - previous_percentage) |>
  mutate(party = "THIRD")

```

`summarise()` has grouped output by 'year', 'state', 'county_name', 'office', 'candidate'. You can override using the `.groups` argument.
 Adding missing grouping variables: `office`, `candidate`

```

plot_2 <- pa_election_df_third |>
  filter(!year %in% c(2000, 2020)) |>
  filter(county_name != "CAMERON") |>
  ggplot() +
  geom_line(aes(x = year, y = swing, color = county_name),
    size = 0.75, alpha = 1
  ) +
  gghighlight(
    county_name %in% county_list,
    label_params = list(size = 2),
    unhighlighted_params = list(size = 0.5, alpha = 0.3)
  ) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(
    title = "Third Party Voting Shifts in Pennsylvania Counties (2004-2016)",
    subtitle = "The same counties that experienced a large two-party shift in
    2016 also experienced a comparatively large third party
    vote presence in 2016 than in previous election years.",
    x = "Election Year",
    y = "Third Party Swing (%)"
  )

```

```

    caption = "Data source: Harvard Dataverse"
  ) +
  scale_x_continuous(
    breaks = seq(2004, 2016, by = 4),
    labels = c(
      "2004\n(Bush)", "2008\n(Obama)",
      "2012\n(Obama)", "2016\n(Trump)"
    ),
    limits = c(2004, 2016),
    expand = expansion(mult = c(0, 0.05))
  ) +
  theme_classic() +
  theme(
    plot.title = element_text(
      margin = margin(b = 5),
      hjust = 0.5,
      face = "bold",
      size = 11.5
    ),
    plot.subtitle = element_text(hjust = 0.5, size = 9.5),
    plot.caption = element_text(size = 6),
    plot.margin = margin(l = 60),
    legend.position = "none",
    panel.grid.major.y = element_line(color = "gray", size = 0.15),
    panel.grid.minor.y = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank(),
    axis.text.x = element_text(color = c("red", "blue", "blue", "red")),
    axis.title = element_text(size = 7.5)
  )

```

Warning: Tried to calculate with group_by(), but the calculation failed.
Falling back to ungrouped filter operation...

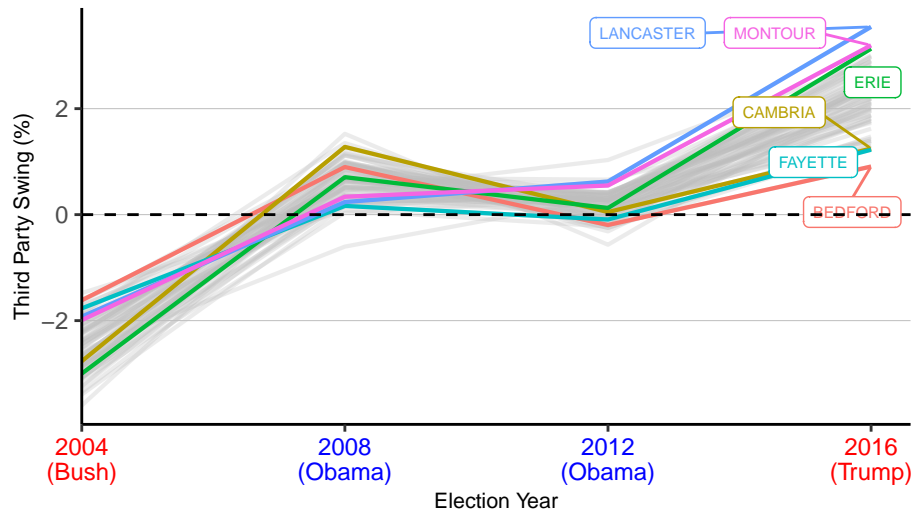
label_key: county_name

Warning: Vectorized input to `element_text()` is not officially supported.
i Results may be unexpected or may change in future versions of ggplot2.

plot_2

Third Party Voting Shifts in Pennsylvania Counties (2004–2

The same counties that experienced a large two-party shift in 2016 also experienced a comparatively large third party vote presence in 2016 than in previous election years.



Data source: Harvard Dataverse

```
avg_coords <-  
  map_data("county") |>  
  filter(region == "pennsylvania") |>  
  mutate(subregion = toupper(subregion)) |>  
  group_by(subregion) |>  
  summarize(  
    avg_long = mean(long),  
    avg_lat = mean(lat)  
  )  
  
pa_election_df_rep_rank <- pa_election_df_rep |>  
  mutate(rank = rank(-inc_rep_vote_share))  
  
joined <- avg_coords |>  
  left_join(pa_election_df_rep_rank, by = c("subregion" = "county_name"))
```

```
library(ggplot2)  
pa_county_borders <- map_data("county") |>  
  mutate(subregion = toupper(subregion)) |>  
  filter(region == "pennsylvania") |>
```

```

left_join(pa_election_df_rep_rank, by = c("subregion" = "county_name")) |>
left_join(avg_coords, by = c("subregion" = "subregion"))

pa_cities <- data.frame(
  city = c("Philadelphia", "Pittsburgh", "Harrisburg", "State College"),
  lat = c(39.9526, 40.4406, 40.2732, 40.7934),
  long = c(-75.1652, -79.9959, -76.8841, -77.8612)
)

plot_3 <- pa_county_borders |>
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = inc_rep_vote_share),
    color = "gray27", size = 0.25
  ) +
  geom_point(
    data = pa_cities,
    aes(x = long, y = lat),
    color = "black",
    size = 0.25
  ) +
  geom_text(
    data = pa_cities, aes(
      x = long,
      y = lat,
      label = city
    ), size = 2,
    nudge_x = 0.1,
    nudge_y = 0.1,
    fontface = "bold"
  ) +
  scale_fill_gradient2(
    low = "blue", mid = "white", high = "red", midpoint = 0,
    limits = c(-15, 15)
  ) +
  theme_void() +
  coord_map("polyconic") +
  labs(
    fill = "Increase in Republican Vote Share\nfrom 2012 to 2016 (%)",
    title = "Mapping Republican Vote Share Increase from 2012 to 2016",
    subtitle = "Republican vote share increased primarily in
      non-metropolitan counties and, on average, at a higher
      magnitude than did Democratic vote share anywhere. The Pocono

```

```

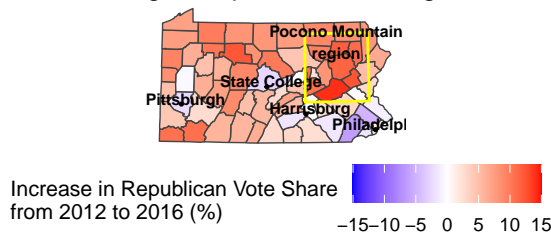
    Mountain region experienced the highest increase.",
    caption = "Data source: Harvard Dataverse"
) +
annotate("rect",
  xmin = -75.3, xmax = -76.9, ymin = 40.5, ymax = 41.8,
  alpha = 0.15, fill = NA, color = "yellow"
) +
annotate("text",
  x = -76.1, y = 41.63, label = "Pocono Mountain\nregion",
  fontface = "bold", size = 2
) +
theme_void() +
theme(
  axis.ticks = element_blank(),
  axis.text.x = element_blank(),
  axis.text.y = element_blank(),
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
  plot.title = element_text(
    margin = margin(b = 8, r = 100, l = 100),
    hjust = 0.5, face = "bold", size = 11.5
  ),
  plot.subtitle = element_text(
    margin = margin(r = 90, l = 100),
    hjust = 0.5, size = 9.5
  ),
  plot.caption = element_text(
    margin = margin(r = 125, l = 500),
    hjust = 0.5, size = 6
  ),
  plot.margin = margin(r = 175, l = 125),
  legend.position = "bottom",
  legend.justification = "center",
  legend.title = element_text(size = 8),
  legend.text = element_text(size = 7),
  legend.key.size = unit(0.5, "cm")
)

```

plot_3

Mapping Republican Vote Share Increase from 2012 to 2016

Republican vote share increased primarily in non-metropolitan counties and, on average, at a higher magnitude than did Democratic vote share anywhere. The Pocono Mountain region experienced the highest increase.



Data source: Harvard Dataver

```
library(patchwork)

combined_plot <- (plot_1 / plot_spacer() / plot_2 / plot_spacer() / plot_3) +
  plot_annotation(
    title = "Breaking Down the Blue Wall in Pennsylvania",
    subtitle = paste(
      "Visualizing the factors that led to a",
      "Republican victory in the state in the 2016 presidential",
      "election"
    ),
    theme = theme(
      plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
      plot.subtitle = element_text(
        margin = margin(b = 20),
        hjust = 0.65, size = 10
      )
    )
  ) +
  plot_layout(heights = c(1, 0.1, 1, 0.1, 1))

ggsave("infographic.pdf",
  plot = combined_plot, device = pdf(),
  width = 10, height = 12.5, units = "in")
```

)