Telling a Story with Data: Economic Indicators by State

Loading, Cleaning, Visualizing, Mapping, and Fine Tuning a story.

Feb 16, 2024

Introduction

Turning Data Analysis into Actionable Insights

- Data storytelling bridges the gap between data analysis and narrative.
- Data analysis should "tell a story". You are the storyteller.

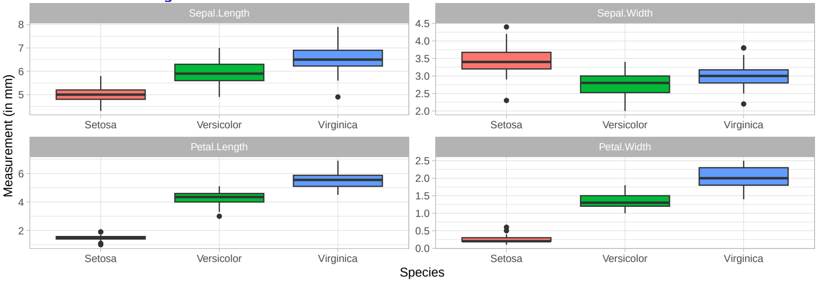
Why this matters.

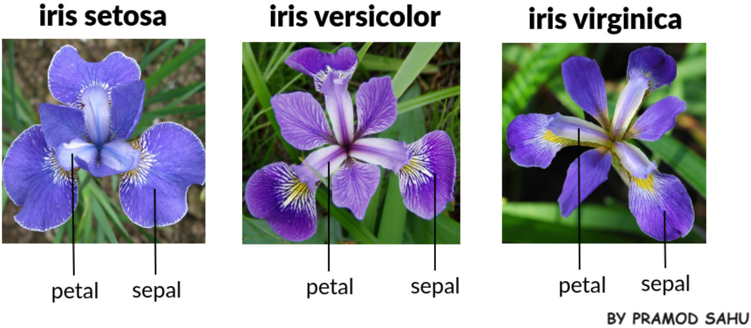
- You, as narrator, place data in the context of a broader goal.
- Through visualization accompanied by writing, you can break down data and its implications for a general audience.

Starting with the most basic of examples: The Iris dataset

```
iris_long <- melt(iris, id.vars = "Species")
iris_long$value <- as.numeric(iris_long$value)
iris_long$Species <- str_to_title(iris_long$Species)

# Plotting with ggplot2
boxplot <- ggplot(iris_long, aes(x = Species, y = value, fi
  geom_boxplot() +
  facet_wrap(~variable, scales = "free") +
  theme_light() +
  labs(y = "Measurement (in mm)", x = "Species") +
  theme(legend.position = "none")
```

What can I say about this?





What do I want to tell a story about?

Economic Indicators by State

- We will explore state-level economic indicators and use some techniques to tell a story about economic diversity across US states.

Setting it up.

Accessing Census Data

tidycensus provides access to a wealth of data from the U.S. Census Bureau, including state-level economic indicators.

```
#census_api_key("f173e6c13a1084bbbd064ef862f1dcd4874697ce")
options(tigris_use_cache = TRUE)
```

Picking the right variables to analyze

- I want to tell a story about economic diversity across US states.
- A standard measure economists use to examine standard of living. is “median household income”
- Why median?
  - Because it tells us how the middle person is doing. It’s unbiased by extreme values.

Fetching State-Level Data

- We'll fetch median household income by state in 2022, a key economic indicator.
- `load_variables()` allows us to view all the codenames for census variables.

```
income_data <- get_acs(geography = "state",
                       variables = "B19013_001",
                       year = 2022,
                       survey = "acs5")
```

Taking a peek.

- State, variable name, estimate, margin of error.

```
head(income_data, 4)
```

```
## # A tibble: 4 x 5
##   GEOID NAME     variable
##   <chr> <chr>    <chr>
## 1 01    Alabama  B19013_~
## 2 02    Alaska   B19013_~
## 3 04    Arizona  B19013_~
## 4 05    Arkansas B19013_~
## # i 2 more variables:
## #   estimate <dbl>,
## #   moe <dbl>
```

Preparing the Data

- The data is already long!
- Let's just fix some names.

```
income_data <- income_data %>%
  rename(state = NAME, median_income = estimate)
```

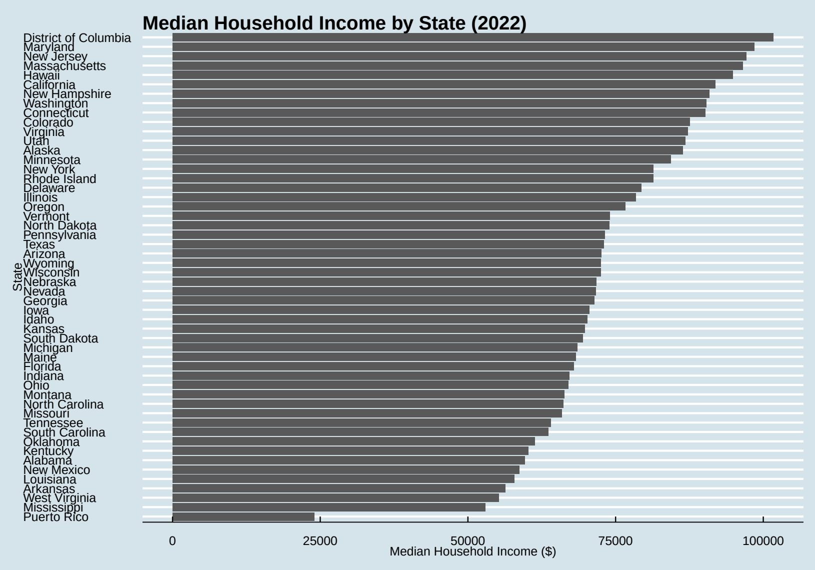
Visualization with ggplot2

Economic Indicator Trends

- We will construct an ordered barplot.
  - First, we treat x as state, ordered by income.
  - y is income.
  - Then we use `coord_flip()` to turn it on its side so the state names fit horizontally.

```
bars <- ggplot(income_data,
 aes(x = reorder(state, median_income), y = median_income))
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Median Household Income by State (2022)",
       x = "State",
       y = "Median Household Income ($)") +
  theme_economist()
```

Tell me what you see.



Mapping Economic Indicators

Bringing in a map.

- `map("state")` from `library(maps)` allows us to read in a map of the US
- `st_as_sf` converts it to a "simple features" object in R, which is our go-to for combining maps with real data.

```
states <- st_as_sf(map(database = "state", fill=TRUE))
```

Merging the map with our income data.

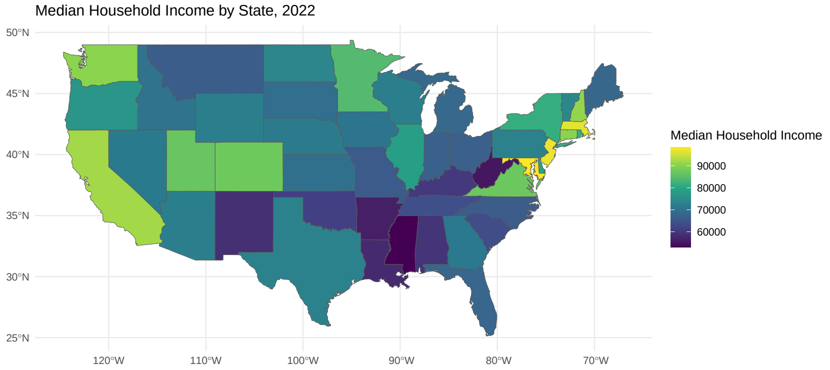
- I want to use `left_join()`, which is like merge, but retains all rows in the first named object (x), without needing them all to be in the second named object (y).
- It turns out we have two problems:
  - State names are stored as `ID` in `states`, and as `state` in `income_data`
  - `states$ID` uses lowercase naming, while `income_data$states` has proper capitalization.
- I capitalize using `str_to_title()` and I `left_join()` with `by = c()`

```
states$ID <- str_to_title(states$ID)
income_data_sf <- left_join(states,
                            income_data, by = c("ID" = "state
```

Making the map object

```
map <- ggplot(income_data_sf) +
  geom_sf(aes(fill = median_income)) +
  scale_fill_viridis_c() +
  labs(title = "Median Household Income by State, 2022",
       fill = "Median Household Income") +
  theme_minimal()
```

Tell me what you see.

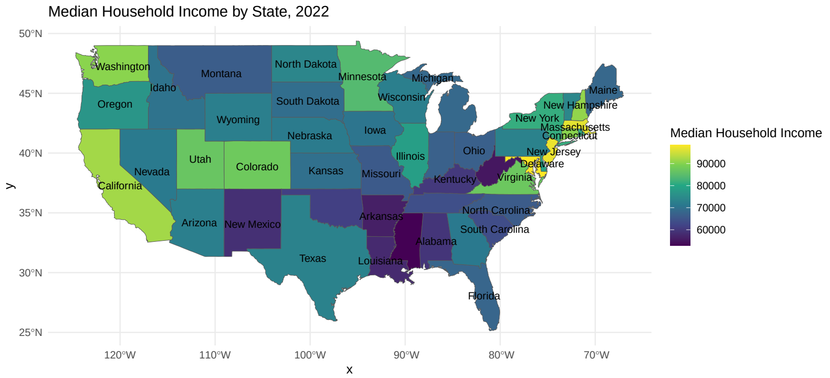


For the geographically challenged (me).

```
map_names <- map + geom_sf_text(aes(label = ID),
                                size = 3,
                                color = "black",
                                check_overlap = TRUE)
```

What do you see?

```
map_names
```



Fine-tuning our story.

What about cost of living?

- People make a lot of money in Maryland, but higher income means higher willingness to pay for goods and services.
  - Higher income –> higher nominal prices.
- Regional price parities measure the differences in price levels across states for a given year and are expressed as a percentage of the overall national price level.

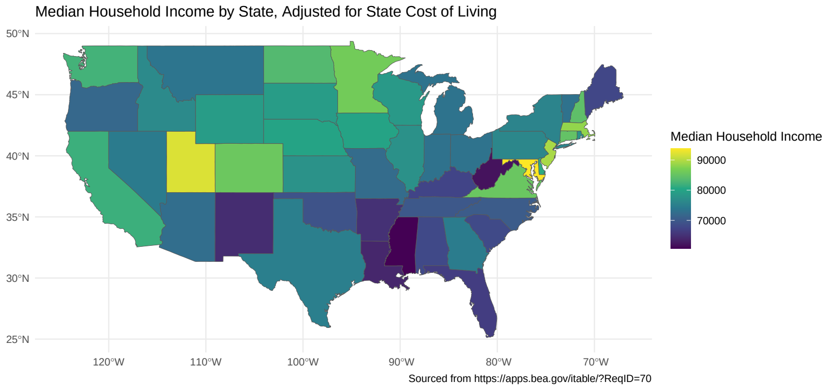
Bringing in that data.

```
rpp <- read.csv("Table.csv")

income_data_sf_index <- left_join(income_data_sf,
                                  rpp,
                                  by = c("ID" = "GeoName"))


income_data_sf_index <- income_data_sf_index %>%
  mutate(median_income_adj = median_income*100/RPP)
```

Making the map object

```
map_adj <- ggplot(income_data_sf_index) +
  geom_sf(aes(fill = median_income_adj)) +
  scale_fill_viridis_c() +
  labs(title = "Median Household Income by State, Adjusted
       fill = "Median Household Income", caption = "Sourced
  theme_minimal()
```

What do you see?



Another story: one of how income has changed in these states from 2019 to 2022.

Fetching Data for 2019 and 2022

```
income_data_2019 <- get_acs(geography = "state", variables
income_data_2022 <- get_acs(geography = "state", variables

# Renaming for clarity
income_data_2019 <- income_data_2019 %>% rename(state = NAM
income_data_2022 <- income_data_2022 %>% rename(state = NAM
```

Calculating Percent Change

```r
# Joining the datasets
income_change <- left_join(income_data_2019, income_data_20

# Calculating percent change
income_change <- income_change %>%
  mutate(pct_change_real = ((median_income_2022 - median_in
```

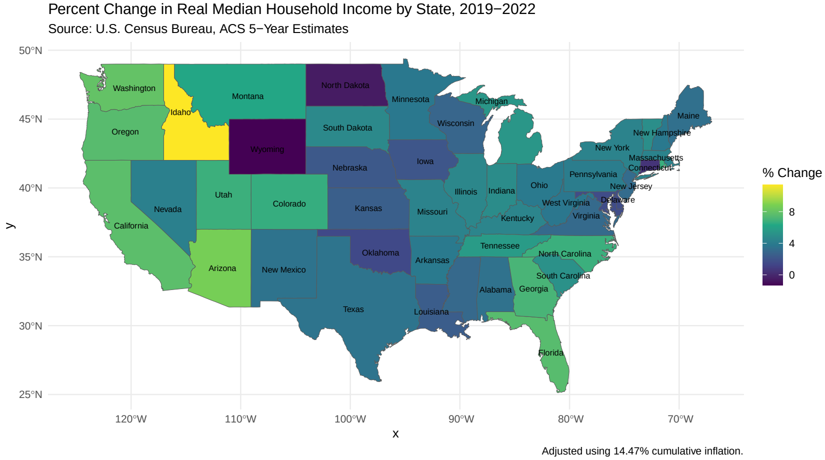
Preparing the Map

```
income_change_map <- left_join(states, income_change, by =
```

Mapping Percent Change

```
pct_change_map <- ggplot(income_change_map) +
  geom_sf(aes(fill = pct_change_real)) +
  scale_fill_viridis_c(name = "% Change") +
  labs(title = "Percent Change in Real Median Household Inc
       subtitle = "Source: U.S. Census Bureau, ACS 5-Year E
       caption = "Adjusted using 14.47% cumulative inflatio
  theme_minimal() +
  geom_sf_text(aes(label = ID), size = 2.5, check_overlap =
```

Now what is our story?



Very advanced: Digging into Idaho.

Disclaimer

- I am showing you this because it's cool.
- I will never ask you to replicate something this complicated on your own.
- But, hopefully it gives you some more ideas about how we can continue to:
  - Investigate further.
  - Develop our story.
  - Do cool things with R.

Pulling Idaho's County Income Data

```
income_data_idaho_2019 <- get_acs(geography = "county",
                                  variables = "B19013_001",
                                  year = 2019,
                                  state = "ID",
                                  geometry = TRUE,
                                  survey = "acs5")

income_data_idaho_2022 <- get_acs(geography = "county",
                                  variables = "B19013_001",
                                  year = 2022,
                                  state = "ID",
                                  geometry = TRUE,
                                  survey = "acs5")

# Renaming for clarity
income_data_idaho_2019 <- income_data_idaho_2019 %>%
  rename(median_income_2019 = estimate)
```

Merging the Idaho years.

```
# Joining the datasets
income_data_idaho_2022_non_spatial <- income_data_idaho_202
  select(GEOID, median_income_2022) %>%
  as.data.frame()

# Now, perform the join
income_growth_idaho <- income_data_idaho_2019 %>%
  left_join(income_data_idaho_2022_non_spatial, by = "GEOII

# Calculating percent change and adjusting for inflation
income_growth_idaho <- income_growth_idaho %>%
  mutate(income_growth_adjusted = ((median_income_2022 - me
```

Finally mapping it.

```
idaho <- ggplot(income_growth_idaho) +
  geom_sf(aes(fill = income_growth_adjusted), color = "whit
  labs(title = "Real Median Household Income Growth by Cour
       subtitle = "Inflation-adjusted using a 14.47% Cumula
       caption = "Source: U.S. Census Bureau, ACS 5-Year Es
       fill="Adjusted Income Growth")  + theme_minimal()+
  theme(legend.position = "right")
```

What do we see?