

# Merging Data

Ben Goldstone

10/06/2023

## Import Dataset

```
States9 = read_csv("~/CSVs/States9.csv")

## Rows: 51 Columns: 6

## -- Column specification -----
## Delimiter: ","
## chr (4): State, AvgPrice, PolitalParty, Region
## dbl (1): BorderingStates

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
States9 = States9 %>% mutate_all(~gsub("\xa0","",.)) %>%
  rename_at("PolitalParty",~"PoliticalParty")
States9Truncated = States9[c("State","AvgPrice","Population","PoliticalParty","Region",
                             "BorderingStates")]

States10 = read_csv("~/CSVs/States10.csv")

## Rows: 50 Columns: 6

## -- Column specification -----
## Delimiter: ","
## chr (1): State
## dbl (5): Population, IncomeTax, PropertyTax, UnemploymentRates, MinimumWage

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
States10 = States10 %>% mutate(UnemploymentRates = UnemploymentRates/100)
States10Truncated = States10[c("State","IncomeTax","PropertyTax","UnemploymentRates","MinimumWage")]
AllStates = full_join(States9Truncated,States10Truncated)

## Joining, by = "State"

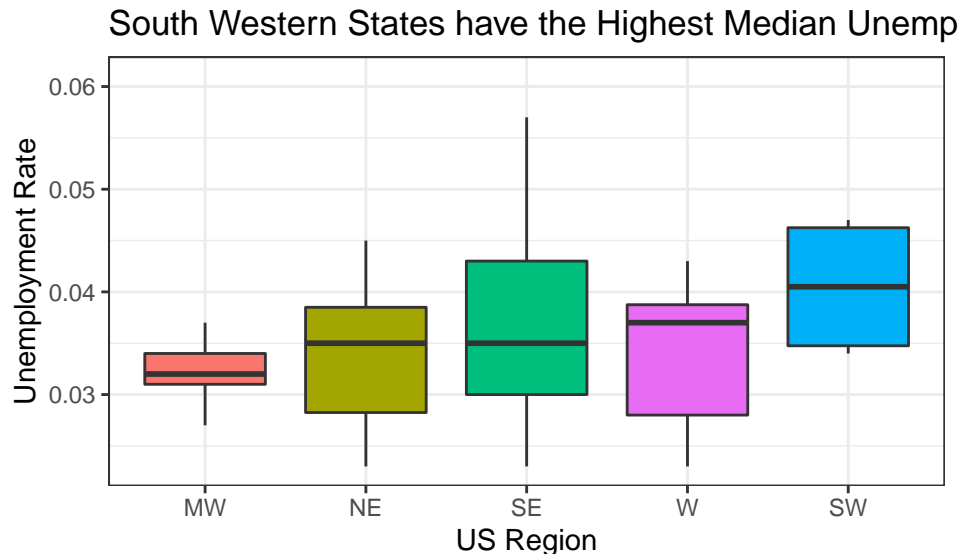
AllStates

## # A tibble: 51 x 10
##   State      AvgPrice Population PoliticalParty Region BorderingStates IncomeTax
##   <chr>      <chr>      <chr>      <chr>      <chr>  <chr>          <dbl>
## 1 Alabama    1.880      4903185    R          SE      4             0.05
## 2 Alaska     2.686      731545    R          W       0             0
```

```
## 3 Arizona 2.563 7278717 R SW 5 0.0454
## 4 Arkansas 1.862 3017825 R SE 6 0.069
## 5 Californ~ 3.175 39512223 D W 3 0.133
## 6 Colorado 2.063 5758736 D W 7 0.0463
## 7 Connecti~ 2.274 3565287 D NE 3 0.0699
## 8 Delaware 2.021 973764 D NE 3 0.066
## 9 Washing~ 2.361 705749 D SE 2 NA
## 10 Florida 2.052 21477737 R SE 2 0
## # ... with 41 more rows, and 3 more variables: PropertyTax <dbl>,
## # UnemploymentRates <dbl>, MinimumWage <dbl>
```

## Unemployment Rate and Regions

```
gf_boxplot(UnemploymentRates~fct_reorder(Region,UnemploymentRates,median), data=na.omit(AllStates),
  outlier.shape = NA, fill=~Region) %>%
  gf_theme(legend.position="none") %>%
  gf_labs(title="South Western States have the Highest Median Unemp",
    x="US Region", y="Unemployment Rate")
```



## Average Populations and Regions

```
# Convert Population to Numeric
AllStates$Population = as.numeric(AllStates$Population)/1000000

regions = AllStates %>% group_by(Region) %>% summarize(AvgPopulation = mean(Population))

gf_boxplot(fct_reorder(Region,Population)~Population,data=AllStates,fill=~Region) %>%
  # Labels
  gf_labs(x="Population(Millions)",y="Region",title="Most Region Populations are Heavily Skewed") +
  # Mean Text
  geom_text(data = regions, aes(x = AvgPopulation, y = Region,
    label = sprintf("Mean Population: %.2f", AvgPopulation), color = Region),
    position = position_dodge(width = 0.8), vjust=-1, hjust=-0.5, size=3) +
  # Set scale to 5
  scale_x_continuous(breaks = seq(0, 40, by = 5))
```

## Most Region Populations are Heavily Skewed

