First Look

Elliot Pickens

October 10, 2018

Load some libraries

- You may need to install some of the libraries below. To install package `tidyverse`, for example, run `install.packages("tidyverse")` in the console

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidyselect)
library(tibble)
library(tidyr)
library(dplyr)
library(ggformula)
```

```
## Loading required package: ggstance
```

```
##
## Attaching package: 'ggstance'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh
```

```
##
## New to ggformula?  Try the tutorials:
##  learnr::run_tutorial("introduction", package = "ggformula")
##  learnr::run_tutorial("refining", package = "ggformula")
```

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(broom)
library(ggthemes)
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(leaps)
library(GGally)
```

```
##
## Attaching package: 'GGally'
```

```
## The following object is masked from 'package:dplyr':
##
##     nasa
```

```
library(aod)
library(readr)
library(readxl)
library(stringr)
#'httr', 'rvest', 'xml2'
```

Load some data

```
registration <- read_xlsx("data/minnesota-voter-registration-by-county-since-2000.xlsx", skip = 1)
elections <- read_xlsx("data/minnesota-election-statistics-1950-to-2016.xlsx", skip = 2, col_names = c(
#dists <- read_csv("data/Minnesota_District_edit.csv")
dists <- read_csv("data/dists_t.csv", skip = 1)
```

```
## Warning: Duplicated column names deduplicated: 'Total population' => 'Total
## population_1' [20], 'Total population' => 'Total population_2' [29], 'Total
## population' => 'Total population_3' [36], 'Total population' => 'Total
## population_4' [43], 'With a disability' => 'With a disability_1' [76],
## 'With a disability' => 'With a disability_2' [78], '65 years and
## over' => '65 years and over_1' [79], 'With a disability' => 'With a
## disability_3' [80], 'Different state' => 'Different state_1' [87],
## 'Civilian labor force' => 'Civilian labor force_1' [96], 'Civilian
## employed population 16 years and over' => 'Civilian employed population
## 16 years and over_1' [112], 'Civilian employed population 16 years
## and over' => 'Civilian employed population 16 years and over_2' [126],
## 'Occupied housing units' => 'Occupied housing units_1' [136], 'Occupied
## housing units' => 'Occupied housing units_2' [141], 'Owner-occupied units'
## => 'Owner-occupied units_1' [158], 'Housing units with a mortgage' =>
```

```
## 'Housing units with a mortgage_1' [161], 'Median (dollars)' => 'Median
## (dollars)_1' [169], 'Housing units without a mortgage' => 'Housing
## units without a mortgage_1' [170], 'Median (dollars)' => 'Median
## (dollars)_2' [177], 'Less than $500' => 'Less than $500_1' [179], '$500
## to $999' => '$500 to $999_1' [180], '$1,000 to $1,499' => '$1,000 to
## $1,499_1' [181], '$1,500 to $1,999' => '$1,500 to $1,999_1' [182],
## '$2,000 to $2,499' => '$2,000 to $2,499_1' [183], '$2,500 to $2,999' =>
## '$2,500 to $2,999_1' [184], '$3,000 or more' => '$3,000 or more_1' [185],
## 'Median (dollars)' => 'Median (dollars)_3' [186], '$100,000 to $149,999'
## => '$100,000 to $149,999_1' [196], 'No health insurance coverage' =>
## 'No health insurance coverage_1' [207], 'With related children of the
## householder under 18 years' => 'With related children of the householder
## under 18 years_1' [212], 'With related children of the householder
## under 5 years only' => 'With related children of the householder
## under 5 years only_1' [213], 'With related children of the householder
## under 18 years' => 'With related children of the householder under
## 18 years_2' [215], 'With related children of the householder under 5
## years only' => 'With related children of the householder under 5 years
## only_2' [216], 'Under 18 years' => 'Under 18 years_1' [218], '18 years
## and over' => '18 years and over_1' [222], '65 years and over' => '65
## years and over_2' [223], 'Construction' => 'Construction_1' [246],
## 'Manufacturing' => 'Manufacturing_1' [247], 'Wholesale trade' =>
## 'Wholesale trade_1' [248], 'Retail trade' => 'Retail trade_1' [249],
## 'Information' => 'Information_1' [251], 'Total for all sectors' => 'Total
## for all sectors_1' [263], 'Agriculture, forestry, fishing and hunting' =>
## 'Agriculture, forestry, fishing and hunting_1' [264], 'Mining, quarrying,
## and oil and gas extraction' => 'Mining, quarrying, and oil and gas
## extraction_1' [265], 'Utilities' => 'Utilities_1' [266], 'Construction'
## => 'Construction_2' [267], 'Manufacturing' => 'Manufacturing_2' [268],
## 'Wholesale trade' => 'Wholesale trade_2' [269], 'Retail trade' => 'Retail
## trade_2' [270], 'Transportation and warehousing' => 'Transportation and
## warehousing_1' [271], 'Information' => 'Information_2' [272], 'Finance and
## insurance' => 'Finance and insurance_1' [273], 'Real estate and rental and
## leasing' => 'Real estate and rental and leasing_1' [274], 'Professional,
## scientific, and technical services' => 'Professional, scientific, and
## technical services_1' [275], 'Management of companies and enterprises'
## => 'Management of companies and enterprises_1' [276], 'Administrative and
## support and waste management and remediation services' => 'Administrative
## and support and waste management and remediation services_1' [277],
## 'Educational services' => 'Educational services_1' [278], 'Health care
## and social assistance' => 'Health care and social assistance_1' [279],
## 'Arts, entertainment, and recreation' => 'Arts, entertainment, and
## recreation_1' [280], 'Accommodation and food services' => 'Accommodation
## and food services_1' [281], 'Other services (except public administration)'
## => 'Other services (except public administration)_1' [282], 'Industries not
## classified' => 'Industries not classified_1' [283], 'Total for all sectors'
## => 'Total for all sectors_2' [284], 'Agriculture, forestry, fishing and
## hunting' => 'Agriculture, forestry, fishing and hunting_2' [285], 'Mining,
## quarrying, and oil and gas extraction' => 'Mining, quarrying, and oil and
## gas extraction_2' [286], 'Utilities' => 'Utilities_2' [287], 'Construction'
## => 'Construction_3' [288], 'Manufacturing' => 'Manufacturing_3' [289],
## 'Wholesale trade' => 'Wholesale trade_3' [290], 'Retail trade' => 'Retail
## trade_3' [291], 'Transportation and warehousing' => 'Transportation and
## warehousing_2' [292], 'Information' => 'Information_3' [293], 'Finance and
```

```
## insurance' => 'Finance and insurance_2' [294], 'Real estate and rental and
## leasing' => 'Real estate and rental and leasing_2' [295], 'Professional,
## scientific, and technical services' => 'Professional, scientific, and
## technical services_2' [296], 'Management of companies and enterprises'
## => 'Management of companies and enterprises_2' [297], 'Administrative and
## support and waste management and remediation services' => 'Administrative
## and support and waste management and remediation services_2' [298],
## 'Educational services' => 'Educational services_2' [299], 'Health care
## and social assistance' => 'Health care and social assistance_2' [300],
## 'Arts, entertainment, and recreation' => 'Arts, entertainment, and
## recreation_2' [301], 'Accommodation and food services' => 'Accommodation
## and food services_2' [302], 'Other services (except public administration)'
## => 'Other services (except public administration)_2' [303], 'Industries not
## classified' => 'Industries not classified_2' [304]
```

```
## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   `Median age (years)` = col_double(),
##   `Unemployment Rate` = col_double(),
##   `Mean travel time to work (minutes)` = col_double(),
##   `Homeowner vacancy rate` = col_double(),
##   `Rental vacancy rate` = col_double(),
##   `Average household size of owner-occupied unit` = col_double(),
##   `Average household size of renter-occupied unit` = col_double(),
##   `All families` = col_double(),
##   `With related children of the householder under 18 years` = col_double(),
##   `With related children of the householder under 5 years only` = col_double(),
##   `Married couple families` = col_double(),
##   `With related children of the householder under 18 years_1` = col_double(),
##   `With related children of the householder under 5 years only_1` = col_double(),
##   `Families with female householder, no husband present` = col_double(),
##   `With related children of the householder under 18 years_2` = col_double(),
##   `With related children of the householder under 5 years only_2` = col_double(),
##   `All people` = col_double(),
##   `Under 18 years_1` = col_double(),
##   `Related children of the householder under 18 years` = col_double(),
##   `Related children of the householder under 5 years` = col_double()
##   # ... with 13 more columns
## )
```

```
## See spec(...) for full column specifications.
```

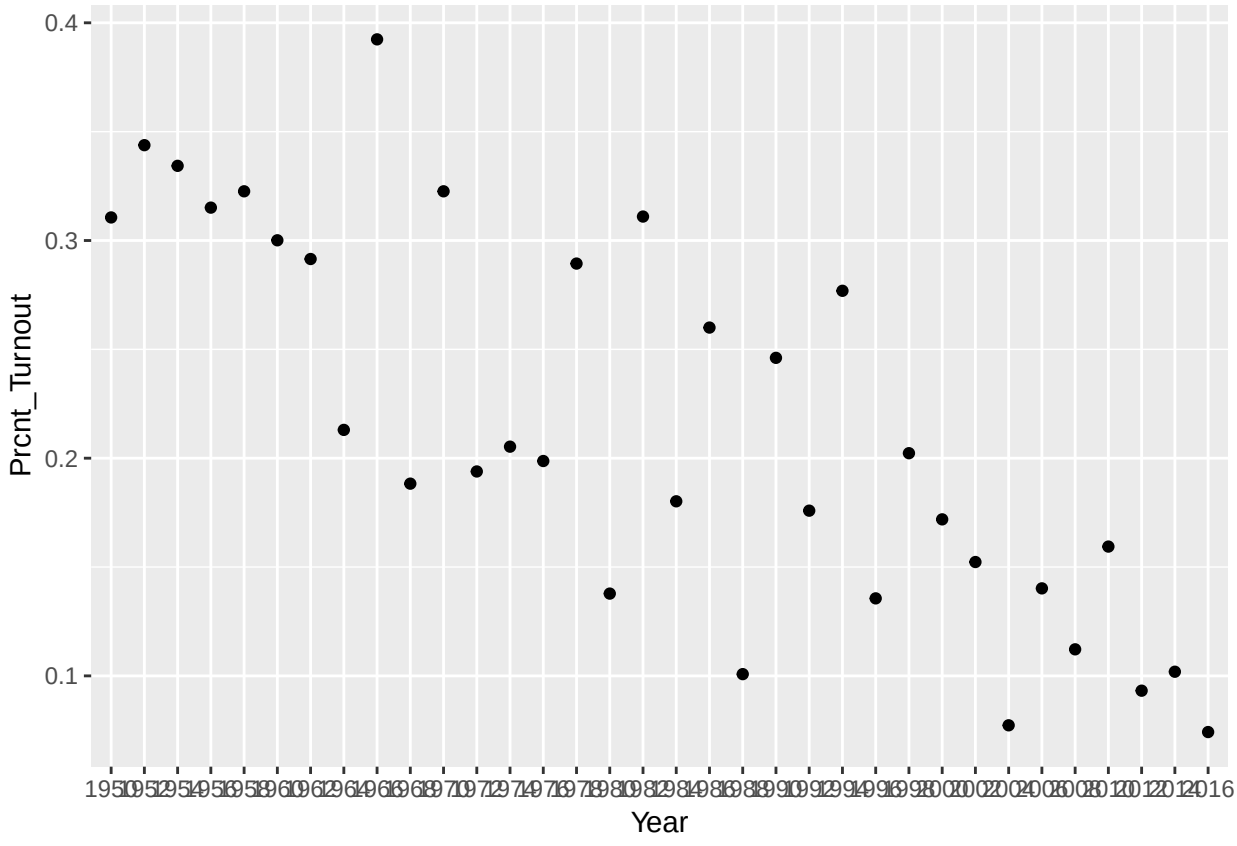
Let's try to clean up this data a little bit

```
elections1 <- elections %>%
  filter(Year <= 2016)

#This ain't right
elect_rm_na <- elections1 %>%
  filter(Prcnt_Turnout == "No Data")


#dist %>%
  #gather(key = var_name, value = value, 2:ncol(dist)) %>%
```

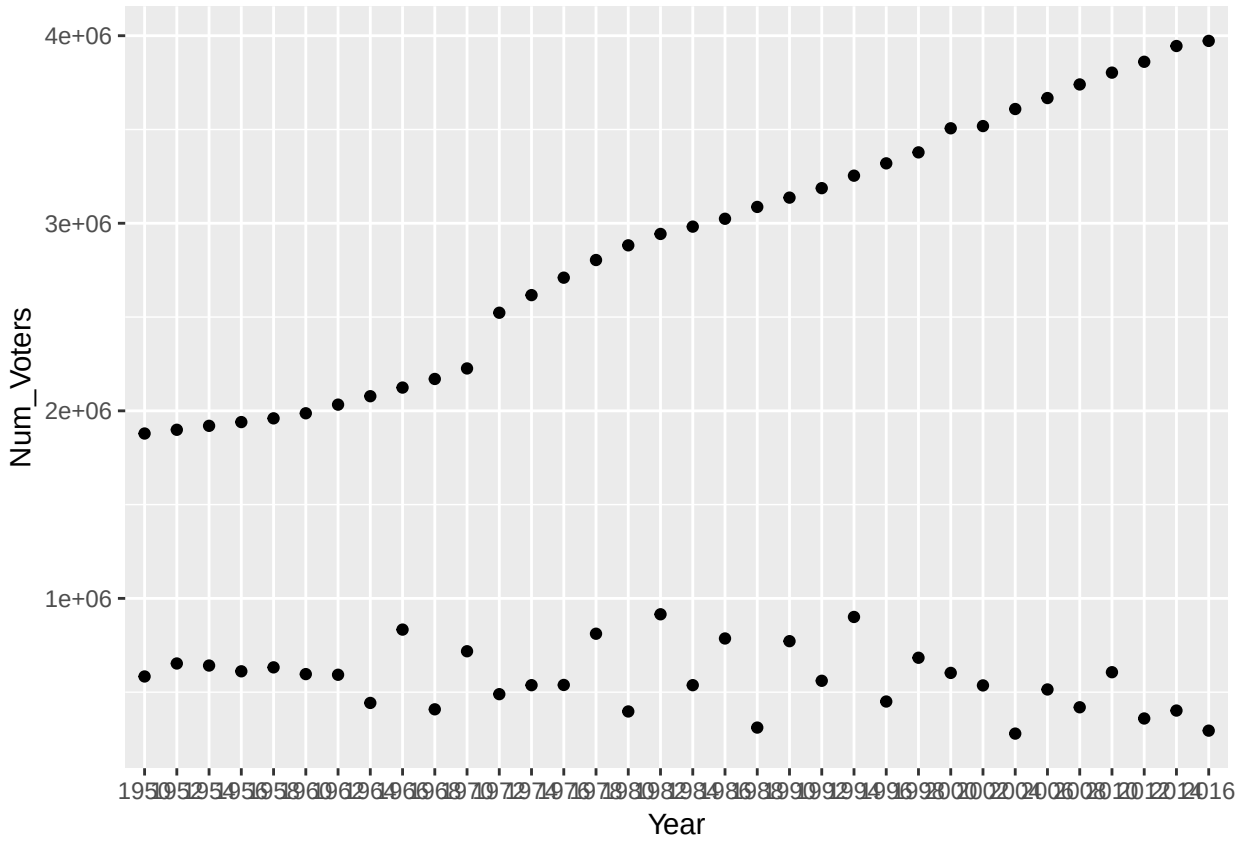
```
  #spread_(key = names(dist)[1], value = 'value')

#dist.df <- as.data.frame(dists)

#t_df <- data.table::transpose(dist.df)
#colnames(t_df) <- rownames(dist.df)
#rownames(t_df) <- colnames(dist.df)

#dists_t <- as_tibble(t_df)
#write_csv(dists_t, "dists_t.csv")
```

Basic Visualization

```
#ggplot(data = elections) +
#  geom_point(mapping = aes(x = Year, y = Prcnt_Turnout))

elections1 %>%
  gf_point(Prcnt_Turnout ~ Year) %>%
    gf_lm(Prcnt_Turnout ~ Year, color = "blue")
```




```
elections1 %>%
  gf_point(Num_Voters ~ Year) %>%
  gf_point(Est_Num_Elg_Voters ~ Year) %>%
  gf_lm(Est_Num_Elg_Voters ~ Year)
```


4e+06
3e+06
2e+06
1e+06
Num_Voters
Year