Fitting an SLR Model

Brooke Walters

2023-05-23

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.2     v purrr   1.0.1
## v tibble  3.2.1     v dplyr   1.1.2
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 0.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
## Warning: package 'stringr' was built under R version 4.2.3
```

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

Load the data

```
# "https://edg.epa.gov/EPADataCommons/public/OA/EPA_SmartLocationDatabase_V3_Jan_2021_Final.csv" # too
walkable <- read.csv("~/gvsu/summer 23/stat 631/data(too big for git)/walkable_2021.csv")
```

Explore the data

```
colnames(walkable)
```

```
##   [1] "OBJECTID"     "GEOID10"      "GEOID20"      "STATEFP"      "COUNTYFP"
##   [6] "TRACTCE"      "BLKGRPCE"     "CSA"          "CSA_Name"     "CBSA"
##  [11] "CBSA_Name"    "CBSA_POP"     "CBSA_EMP"     "CBSA_WRK"     "Ac_Total"
##  [16] "Ac_Water"     "Ac_Land"      "Ac_Unpr"      "TotPop"       "CountHU"
##  [21] "HH"           "P_WrkAge"     "AutoOwn0"     "Pct_AO0"      "AutoOwn1"
##  [26] "Pct_AO1"      "AutoOwn2p"    "Pct_AO2p"     "Workers"      "R_LowWageWk"
##  [31] "R_MedWageWk"  "R_HiWageWk"   "R_PCTLOWWAGE" "TotEmp"       "E5_Ret"
```

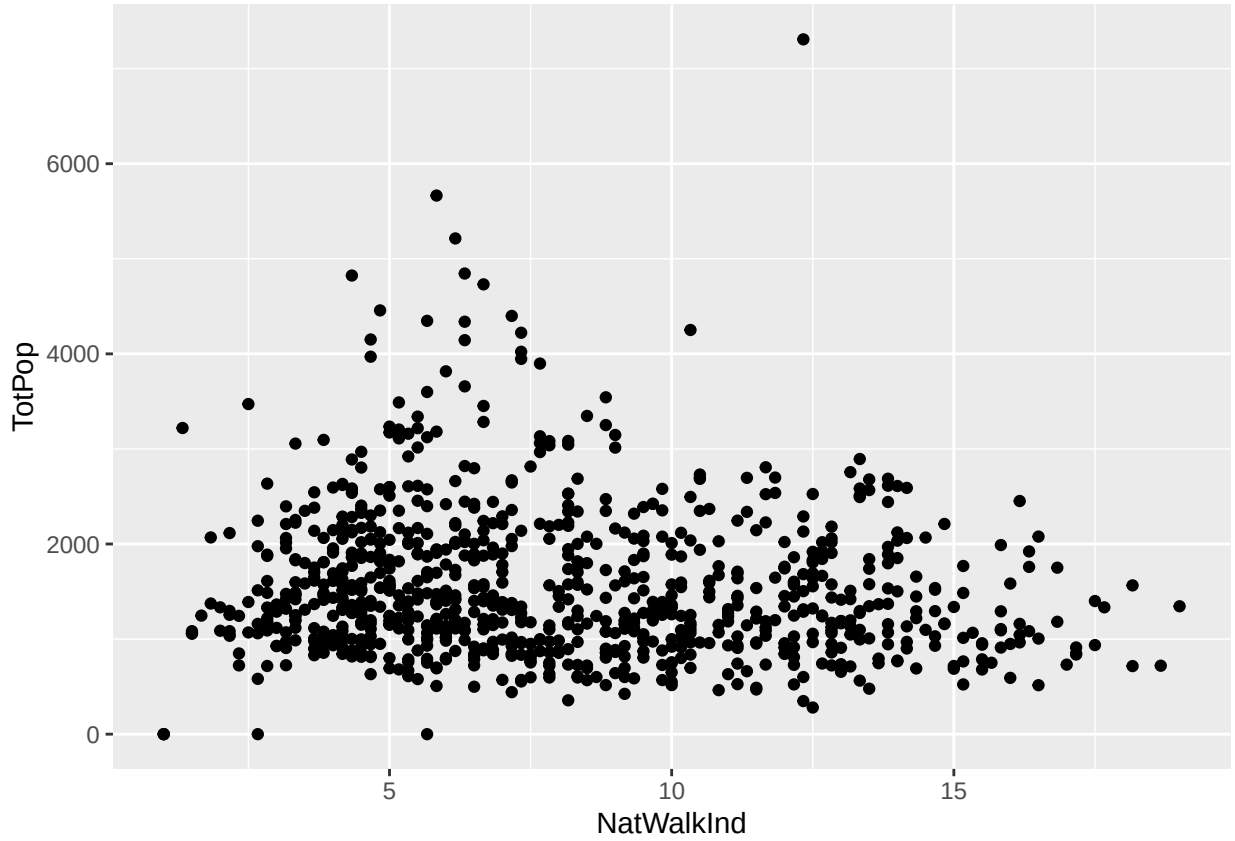
```
##  [36] "E5_Off"       "E5_Ind"       "E5_Svc"       "E5_Ent"       "E8_Ret"
##  [41] "E8_off"       "E8_Ind"       "E8_Svc"       "E8_Ent"       "E8_Ed"
##  [46] "E8_Hlth"      "E8_Pub"       "E_LowWageWk"  "E_MedWageWk"  "E_HiWageWk"
##  [51] "E_PctLowWage" "D1A"          "D1B"          "D1C"          "D1C5_RET"
##  [56] "D1C5_OFF"     "D1C5_IND"     "D1C5_SVC"     "D1C5_ENT"     "D1C8_RET"
##  [61] "D1C8_OFF"     "D1C8_IND"     "D1C8_SVC"     "D1C8_ENT"     "D1C8_ED"
##  [66] "D1C8_HLTH"    "D1C8_PUB"     "D1D"          "D1_FLAG"      "D2A_JPHH"
##  [71] "D2B_E5MIX"    "D2B_E5MIXA"   "D2B_E8MIX"    "D2B_E8MIXA"   "D2A_EPHHM"
##  [76] "D2C_TRPMX1"   "D2C_TRPMX2"   "D2C_TRIPEQ"   "D2R_JOBPOP"   "D2R_WRKEMP"
##  [81] "D2A_WRKEMP"   "D2C_WREMLX"   "D3A"          "D3AAO"        "D3AMM"
##  [86] "D3APO"        "D3B"          "D3BAO"        "D3BMM3"       "D3BMM4"
##  [91] "D3BPO3"       "D3BPO4"       "D4A"          "D4B025"       "D4B050"
##  [96] "D4C"          "D4D"          "D4E"          "D5AR"         "D5AE"
## [101] "D5BR"         "D5BE"         "D5CR"         "D5CRI"        "D5CE"
## [106] "D5CEI"        "D5DR"         "D5DRI"        "D5DE"         "D5DEI"
## [111] "D2A_Ranked"   "D2B_Ranked"   "D3B_Ranked"   "D4A_Ranked"   "NatWalkInd"
## [116] "Shape_Length" "Shape_Area"
```

```
walkit <- walkable %>%
  filter( CSA_Name == 'Grand Rapids-Kentwood-Muskegon, MI') %>%
  select(NatWalkInd, TotPop, CSA_Name, Workers, Ac_Land)

head(walkit)
```

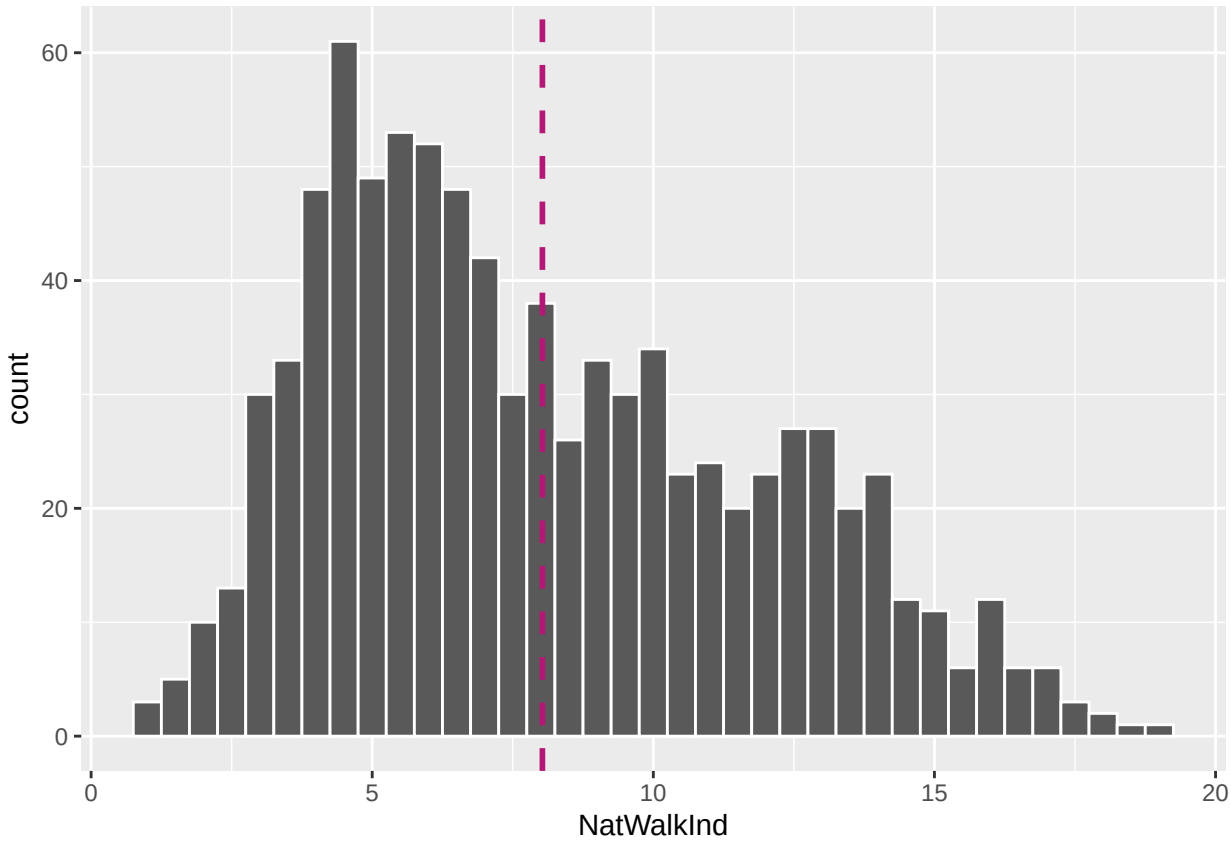
```
##   NatWalkInd TotPop                           CSA_Name Workers  Ac_Land
## 1   7.333333   1224 Grand Rapids-Kentwood-Muskegon, MI     642 788.9070
## 2   4.500000   2019 Grand Rapids-Kentwood-Muskegon, MI    1108 517.8695
## 3   5.000000   1158 Grand Rapids-Kentwood-Muskegon, MI     548 176.8762
## 4   5.333333   2023 Grand Rapids-Kentwood-Muskegon, MI    1002 709.0950
## 5  11.166667   2246 Grand Rapids-Kentwood-Muskegon, MI     982 275.7735
## 6   7.666667    873 Grand Rapids-Kentwood-Muskegon, MI     516 232.0321
```

```
# make some plots
ggplot(walkit, aes(x = NatWalkInd, y = TotPop))+
geom_point()
```

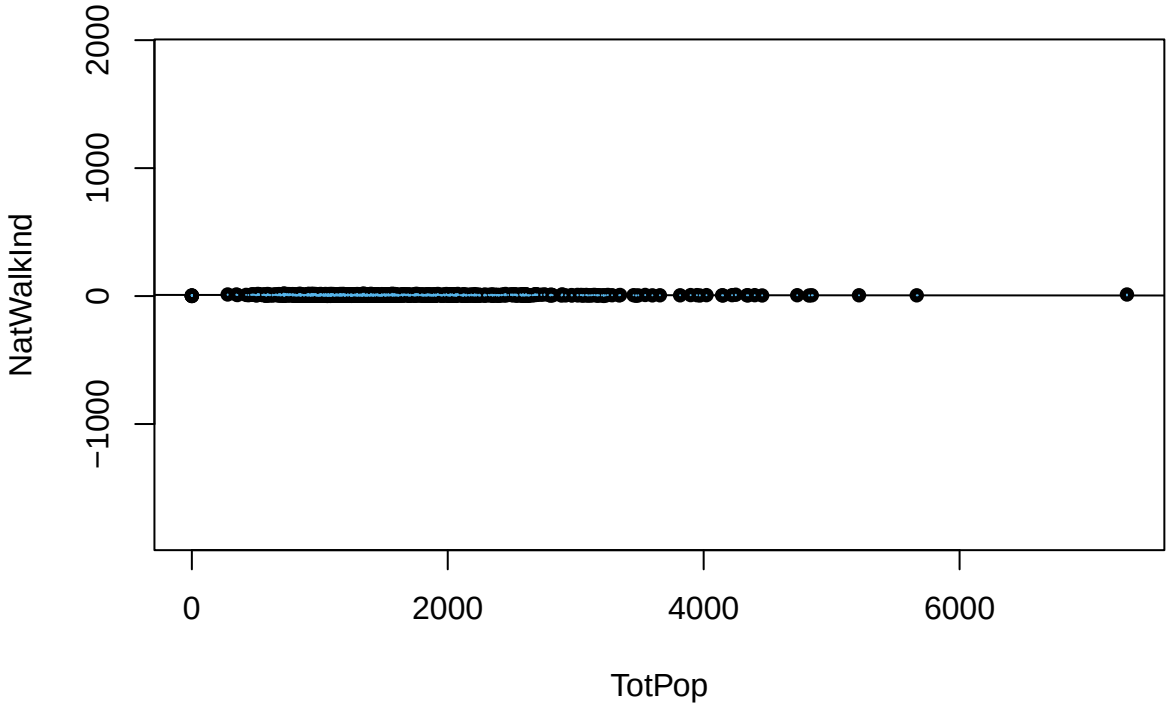



```
# make a histogram
ggplot(walkit, aes(x=NatWalkInd)) +
    geom_histogram(binwidth=.5, colour="white") +
    geom_vline(aes(xintercept=mean(NatWalkInd, na.rm=T)),
               color="#B41876", linetype="dashed", size=1)
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```




```
#for the console
statsr::plot_ss(x = TotPop, y = NatWalkInd, data = walkit)
```




```
## Click two points to make a line.
## Call:
## lm(formula = y ~ x, data = pts)
##
## Coefficients:
## (Intercept)            x
##    8.897806    -0.000556
##
## Sum of Squares:  12629.47
```

```
library(tidymodels)
```

```
## Warning: package 'tidymodels' was built under R version 4.2.3
```

```
## -- Attaching packages --------------------------------------- tidymodels 1.1.0 --
```

```
## v broom        1.0.4     v rsample      1.1.1
## v dials        1.2.0     v tune         1.1.1
## v infer        1.0.4     v workflows    1.1.3
## v modeldata    1.1.0     v workflowsets 1.0.1
## v parsnip      1.1.0     v yardstick    1.2.0
## v recipes      1.0.6
```

```
## Warning: package 'broom' was built under R version 4.2.3
```

```
## Warning: package 'dials' was built under R version 4.2.3

## Warning: package 'infer' was built under R version 4.2.3

## Warning: package 'modeldata' was built under R version 4.2.3

## Warning: package 'parsnip' was built under R version 4.2.3

## Warning: package 'recipes' was built under R version 4.2.3

## Warning: package 'rsample' was built under R version 4.2.3

## Warning: package 'tune' was built under R version 4.2.3

## Warning: package 'workflows' was built under R version 4.2.3

## Warning: package 'workflowsets' was built under R version 4.2.3

## Warning: package 'yardstick' was built under R version 4.2.3

## -- Conflicts --------------------------------------- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Learn how to get started at https://www.tidymodels.org/start/
```

```
# Created a parsnip specification for a linear model

# , "parsnip" refers to a package that provides a consistent interface for working with various modelin

lm_spec <- linear_reg() %>%
  set_mode("regression") %>%
  set_engine("lm")

# Fit our specification to our data

slr_mod <- fit(lm_spec, formula = TotPop ~ NatWalkInd, data = walkit)


# View our model output
tidy(slr_mod)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   1762.      62.2      28.3  4.60e-126
## 2 NatWalkInd     -24.4      7.00     -3.49 5.13e-  4
```