Activity 3 - MLR

```
# install packages
library(tidyverse)
library(tidymodels)
library(ggthemes)
library(GGally)
```

import data data dictionary : https://www.epa.gov/smartgrowth/smart-location-mapping#walkability

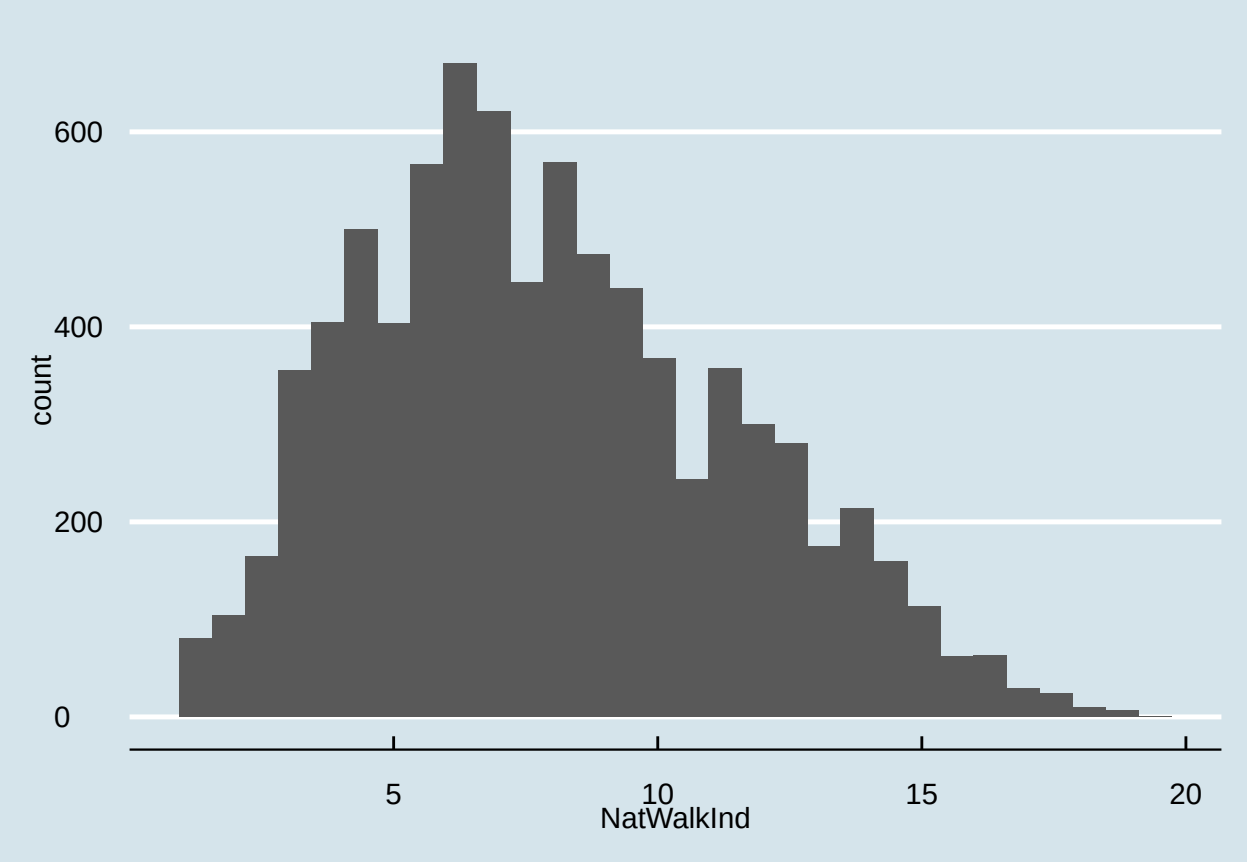
```
# "https://edg.epa.gov/EPADataCommons/public/OA/EPA_SmartLocationDatabase_V3_Jan_2021_Final.csv" # too
# walkable <- read.csv("~/gvsu/summer 23/stat 631/data(too big for git)/walkable_2021.csv")
#
# walkit <- walkable %>%
# filter( STATEFP == '26')
#
# write.csv(walkit, "mi_walkable_2021.csv")
walkit <- read.csv("mi_walkable_2021.csv")
```

```
colnames(walkit)
```

```
##   [1] "X"             "OBJECTID"      "GEOID10"       "GEOID20"       "STATEFP"
##   [6] "COUNTYFP"      "TRACTCE"       "BLKGRPCE"      "CSA"           "CSA_Name"
##  [11] "CBSA"          "CBSA_Name"     "CBSA_POP"      "CBSA_EMP"      "CBSA_WRK"
##  [16] "Ac_Total"      "Ac_Water"      "Ac_Land"       "Ac_Unpr"       "TotPop"
##  [21] "CountHU"       "HH"            "P_WrkAge"      "AutoOwn0"      "Pct_AO0"
##  [26] "AutoOwn1"      "Pct_AO1"       "AutoOwn2p"     "Pct_AO2p"      "Workers"
##  [31] "R_LowWageWk"   "R_MedWageWk"   "R_HiWageWk"    "R_PCTLOWWAGE"  "TotEmp"
##  [36] "E5_Ret"        "E5_Off"        "E5_Ind"        "E5_Svc"        "E5_Ent"
##  [41] "E8_Ret"        "E8_off"        "E8_Ind"        "E8_Svc"        "E8_Ent"
##  [46] "E8_Ed"         "E8_Hlth"       "E8_Pub"        "E_LowWageWk"   "E_MedWageWk"
##  [51] "E_HiWageWk"    "E_PctLowWage"  "D1A"           "D1B"           "D1C"
##  [56] "D1C5_RET"      "D1C5_OFF"      "D1C5_IND"      "D1C5_SVC"      "D1C5_ENT"
##  [61] "D1C8_RET"      "D1C8_OFF"      "D1C8_IND"      "D1C8_SVC"      "D1C8_ENT"
##  [66] "D1C8_ED"       "D1C8_HLTH"     "D1C8_PUB"      "D1D"           "D1_FLAG"
##  [71] "D2A_JPHH"      "D2B_E5MIX"     "D2B_E5MIXA"    "D2B_E8MIX"     "D2B_E8MIXA"
##  [76] "D2A_EPHHM"     "D2C_TRPMX1"    "D2C_TRPMX2"    "D2C_TRIPEQ"    "D2R_JOBPOP"
##  [81] "D2R_WRKEMP"    "D2A_WRKEMP"    "D2C_WREMLX"    "D3A"           "D3AAO"
##  [86] "D3AMM"         "D3APO"         "D3B"           "D3BAO"         "D3BMM3"
##  [91] "D3BMM4"        "D3BPO3"        "D3BPO4"        "D4A"           "D4B025"
##  [96] "D4B050"        "D4C"           "D4D"           "D4E"           "D5AR"
## [101] "D5AE"          "D5BR"          "D5BE"          "D5CR"          "D5CRI"
## [106] "D5CE"          "D5CEI"         "D5DR"          "D5DRI"         "D5DE"
## [111] "D5DEI"         "D2A_Ranked"    "D2B_Ranked"    "D3B_Ranked"    "D4A_Ranked"
## [116] "NatWalkInd"    "Shape_Length"  "Shape_Area"
```

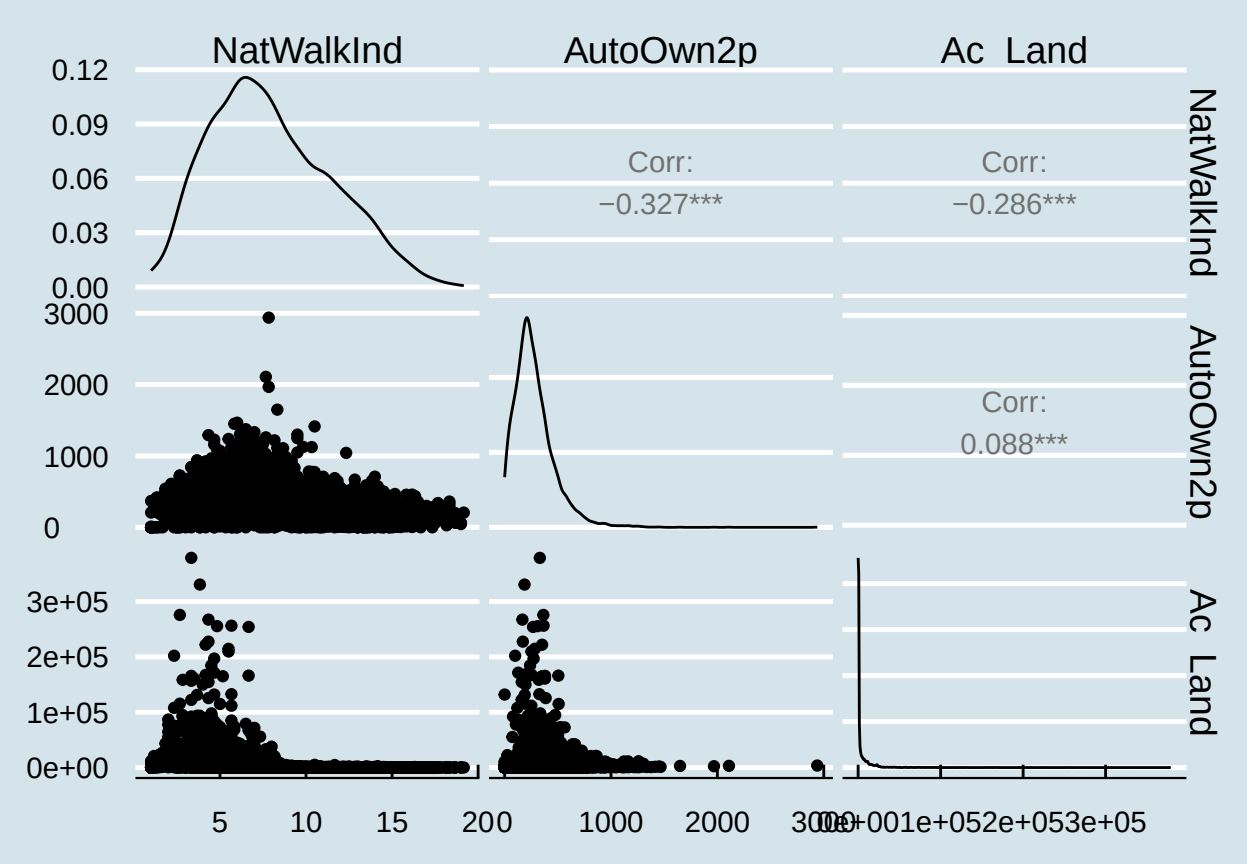
visualize walkability scores accross grand rapids

```
ggplot(walkit, aes(x = NatWalkInd)) +
  geom_histogram(bins = 30) +
   theme_economist()
```



interpretations:

```
  walkit %>%
  select(NatWalkInd, AutoOwn2p, Ac_Land) %>%
  ggpairs() +
  theme_economist()
```




```
# NatWalkInd = Walkability index comprised of weighted sum of the
# ranked values of [D2a_EpHHm] (D2A_Ranked),
# Ac_Land = Total land area (acres)
# R_HiWageWk  = Count of workers earning $3333/month or more (home
# location), 2017

# R_LowWageWk  = Count of workers earning $3333/month or more (home
# location), 2017
# AutoOwn2p = umber of households in CBG that own two or more automobiles, 2018
```

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

lm_spec
```

```
## Linear Regression Model Specification (regression)
##
## Computational engine: lm
```

```
mlr_mod <- lm_spec %>%
fit(NatWalkInd ~ AutoOwn2p + Ac_Land, data = walkit)
tidy(mlr_mod)
```

```
## # A tibble: 3 x 5
```

```
##   term          estimate  std.error statistic   p.value
##   <chr>            <dbl>      <dbl>     <dbl>     <dbl>
## 1 (Intercept)  9.78      0.0627         156.  0
## 2 AutoOwn2p   -0.00562   0.000186       -30.1 1.76e-189
## 3 Ac_Land     -0.0000594 0.00000231     -25.7 1.56e-140
```