Big Data Procurement

Mitchell Beckner

1/23/2021

Objective

The goal of this portion of the project was to obtain a "big data" dataset compiled around United States Census Bureau data. The dataset was to be compiled at the Zip or Zip+4 level, and use 10 year census data as the foundation with additional 5 and/or 1 year census data added if it appeared to be useful. Additional data from outside sources could also be added to the base data if it proved useful.

Methodology

The base data set for this was acquired from the US Census Bureau using their api. R code was written to download the data tables shown below:

10 Year Census Tables

- P2 - Urban and Rural Population
- P3 - Race
- P4 - Hispanic or Latino Origin
- P12 - Sex by Age
- H13 - Household Size

5 Year Census Tables

- B08301 - Means of Transportation to Work
- B15002 - Sex by Education Attainment
- B19013 - Median Household Income
- B23025 - Employment Status
- B28011 - Internet Subscriptions in Household

Non-Census Tables

- USA Cycling Membership data
- NOAA Minimum Temperature by County

Supplemental Files

- HUD USPS Zip Code Crosswalk Files
- US Census Bureau ZCTA5 Crosswalk File

Using the tables above, the desired data was downloaded and duplicate rows were removed. This was done at at the block level of detail for 10 year census tables and at the tract level for 5 year census tables (block level data was not available). The final product will be constructed at the Zip + 4 level once the tract level tables are modified to reflect the higher level of detail. The desired variables were selected and examined. Some preliminary cleaning and processing of variables was performed although more transformation will be necessary prior to model creation. Tables were joined, duplicate columns removed, and a variety of geographic ID fields were added to ease in joining additional data at various levels and from various sources.

Procured Dataset

The resulting dataset is a single 5.39 GB .csv file It contains aggregations of 155 variables from 6,495,834 unique census blocks. Census blocks are the smallest level of geography that basic demographic data can be obtained for and are the building blocks for census block groups which generally contain between 600 and 3,000 people. The data collected covers all 50 States plus the District of Columbia. An overview of the selected variables follows:

- 9 location variables including GEO_ID, census tract #, USPS Zip, ZCTA5, and county fips
- Total, urban, and rural population
- 8 race and ethnicity variables
- 48 gender and age categorizations
- 10 Household variables including median income, size and internet access
- 35 Education level categories
- 7 Employment variables

The variable sets below are more directly related to the target market, the bicycle industry:

- 7 variables related to commuting
- 27 variables on bicycle racing license holders
- Minimum average outdoor temperature by county

Appendix A

Detailed Data Features

Figure 1: Cleaned Data Structure pt.1

Figure 2: Cleaned Data Structure pt.2

```
ASP1.R ×    Block_level_cleaning.R ×    Census_API_import_acs.R ×    USAC_to_block_conversion.R

    1    setwd("C:/Users/mitch/OneDrive/ANLT-530")
    2
    3    library(tidyverse)
    4    library(data.table)
    5
192:1    # (Untitled)

Console    Terminal ×    Jobs ×

C:/Users/mitch/OneDrive/ANLT-530/
$ Female_12th_grade_no_diploma                           <dbl> 0.282525608, 1.
$ Female_High_school_graduate_includes_equivalency       <dbl> 9.1349946, 34.8
$ Female_Some_college_less_than_1_year                   <dbl> 2.14248586, 8.1
$ Female_Some_college_1_or_more_years_no_degree          <dbl> 4.70876013, 17.
$ Female_Associates_degree                               <dbl> 4.82647913, 18.
$ Female_Bachelors_degree                                <dbl> 1.87173215, 7.1
$ Female_Masters_degree                                  <dbl> 2.48387097, 9.4
$ Female_Professional_school_degree                      <dbl> 0, 0, 0, 0, 0,
$ Female_Doctorate_degree                                <dbl> 0, 0, 0, 0, 0,
$ Median_household_income_in_2019_dollars                <dbl> 75194, 75194, 7
$ Households_with_no_Internet_access                     <dbl> 4.2258065, 17.0
$ Total_Employment_Pop                                   <dbl> 65.7225195, 250
$ In_labor_force                                         <dbl> 43.4147684, 165
$ In_labor_force_Civilian_labor_force                    <dbl> 43.2146461, 165
$ In_labor_force_Civilian_labor_force_Employed           <dbl> 40.6836875, 155
$ In_labor_force_Civilian_labor_force_Unemployed         <dbl> 2.53095857, 9.6
$ In_labor_force_Armed_Forces                            <dbl> 0.200122305, 0.
$ Not_in_labor_force                                     <dbl> 22.3077511, 85.
$ Public_transportation_excluding_taxicab                <dbl> 0.153034704, 0.
$ Taxicab                                                <dbl> 0, 0, 0, 0, 0,
$ Motorcycle                                             <dbl> 0, 0, 0, 0, 0,
$ Bicycle                                                <dbl> 0, 0, 0, 0, 0,
$ Walked                                                 <dbl> 0.612138817, 2.
$ Other_means                                            <dbl> 1.11833053, 4.2
$ Worked_from_home                                       <dbl> 1.04769913, 4.0
$ Racers                                                 <dbl> NA, NA, NA, NA,
$ Racers_Male                                            <dbl> NA, NA, NA, NA,
$ Racers_Female                                          <dbl> NA, NA, NA, NA,
$ Racers_Junior                                          <dbl> NA, NA, NA, NA,
$ Racers_Elite                                           <dbl> NA, NA, NA, NA,
$ Racers_Master                                          <dbl> NA, NA, NA, NA,
$ Racers_Pro_Road                                        <dbl> NA, NA, NA, NA,
$ Racers_Pro_MTB                                         <dbl> NA, NA, NA, NA,
$ Racers_Pro_Track                                       <dbl> NA, NA, NA, NA,
$ Racers_Pro_Cross                                       <dbl> NA, NA, NA, NA,
$ Racers_Mid_road                                        <dbl> NA, NA, NA, NA,
$ Racers_Mid_MTB                                         <dbl> NA, NA, NA, NA,
$ Racers_Mid_Track                                       <dbl> NA, NA, NA, NA,
$ Racers_Mid_Cross                                       <dbl> NA, NA, NA, NA,
$ Male_Racers                                            <dbl> NA, NA, NA, NA,
$ Female_Racers                                          <dbl> NA, NA, NA, NA,
$ Junior_Racers                                          <dbl> NA, NA, NA, NA,
$ Elite_Racers                                           <dbl> NA, NA, NA, NA,
$ Master_Racers                                          <dbl> NA, NA, NA, NA,
$ Pro_Road_Racers                                        <dbl> NA, NA, NA, NA,
$ Pro_MTB_Racers                                         <dbl> NA, NA, NA, NA,
$ Pro_Track_Racers                                       <dbl> NA, NA, NA, NA,
$ Pro_Cross_Racers                                       <dbl> NA, NA, NA, NA,
$ Mid_Road_Racers                                        <dbl> NA, NA, NA, NA,
$ Mid_MTB_Racers                                         <dbl> NA, NA, NA, NA,
$ Mid_Track_Racers                                       <dbl> NA, NA, NA, NA,
$ Mid_Cross_Racers                                       <dbl> NA, NA, NA, NA,
$ Min_Temp                                               <dbl> 29.1, 29.1, 29.
> summary(df_final)
```
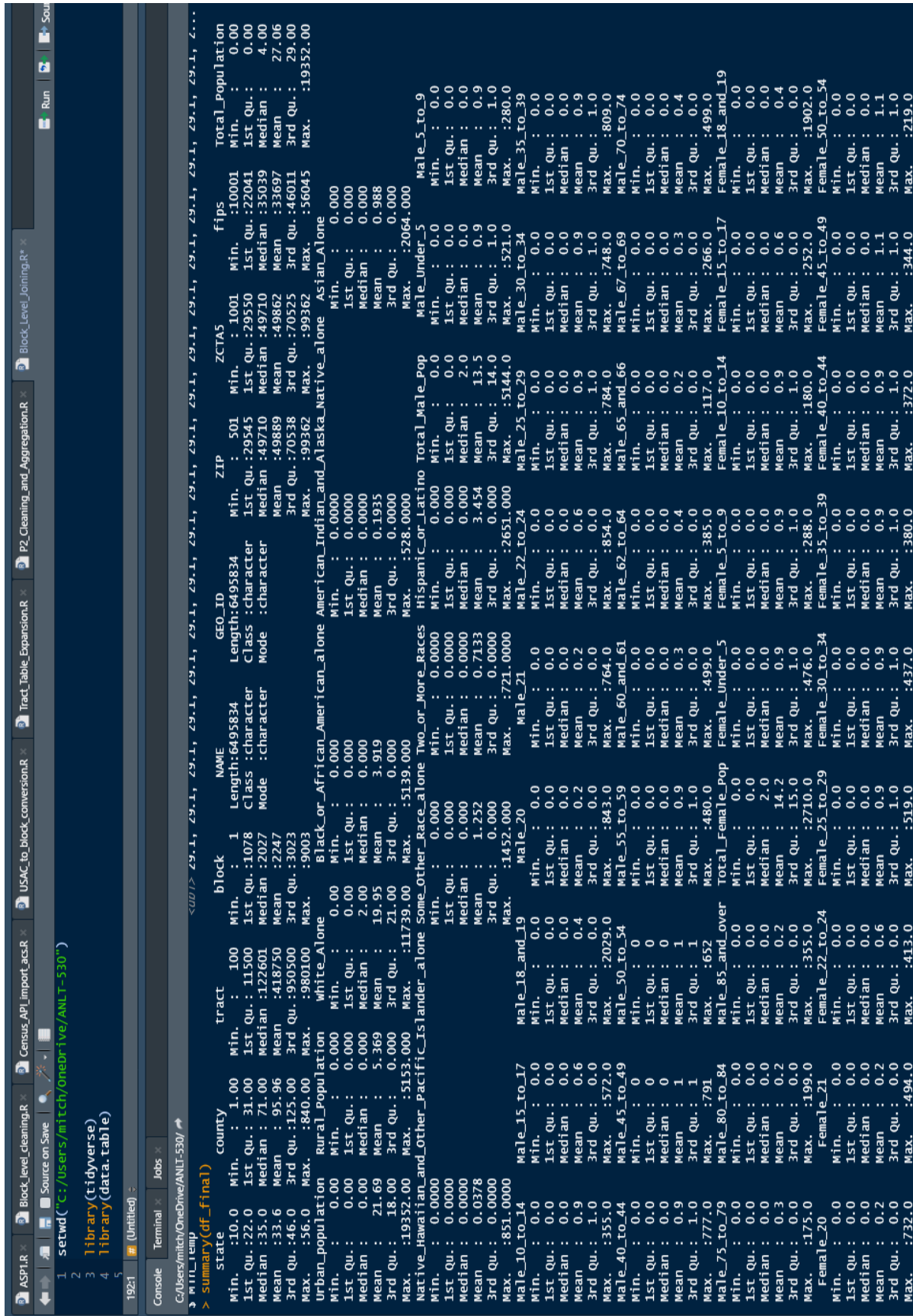
Figure 3: Cleaned Data Structure pt.3

6

```r
setwd("C:/Users/mitch/OneDrive/ANLT-530")

library(tidyverse)
library(data.table)
```

```
> summary(df_final)
     state           county            tract             block           NAME             GEO_ID             ZIP            ZCTA5            fips        Total_Population
 Min.   :10.0    Min.   :  1.00    Min.   :   100    Min.   :   1    Length:6495834     Length:6495834    Min.   :  501    Min.   : 1001    Min.   :10001    Min.   :    0.00
 1st Qu.:22.0    1st Qu.: 31.00    1st Qu.: 11500    1st Qu.:1078    Class :character   Class :character  1st Qu.:29545    1st Qu.:29550    1st Qu.:22041    1st Qu.:    0.00
 Median :35.0    Median : 71.00    Median :122601    Median :2027    Mode  :character   Mode  :character  Median :49710    Median :49862    Median :35039    Median :    4.00
 Mean   :33.6    Mean   : 95.96    Mean   :418750    Mean   :2247                                          Mean   :49889    Mean   :49889    Mean   :33697    Mean   :   27.06
 3rd Qu.:46.0    3rd Qu.:125.00    3rd Qu.:950500    3rd Qu.:3023                                          3rd Qu.:70538    3rd Qu.:70525    3rd Qu.:46011    3rd Qu.:   29.00
 Max.   :56.0    Max.   :840.00    Max.   :980100    Max.   :9003                                          Max.   :99362    Max.   :99362    Max.   :56045    Max.   :19352.00

 urban_population   Rural_Population    White_Alone      Black_or_African_American_alone  American_Indian_and_Alaska_Native_alone  Asian_Alone
 Min.   :    0.00   Min.   :   0.000   Min.   :    0.00   Min.   :   0.000                 Min.   :  0.0000                         Min.   :   0.000
 1st Qu.:    0.00   1st Qu.:   0.000   1st Qu.:    0.00   1st Qu.:   0.000                 1st Qu.:  0.0000                         1st Qu.:   0.000
 Median :    0.00   Median :   0.000   Median :    2.00   Median :   0.000                 Median :  0.0000                         Median :   0.000
 Mean   :   21.69   Mean   :   5.369   Mean   :   19.95   Mean   :   3.919                 Mean   :  0.1935                         Mean   :   3.919
 3rd Qu.:   18.00   3rd Qu.:   0.000   3rd Qu.:   21.00   3rd Qu.:   0.000                 3rd Qu.:  0.0000                         3rd Qu.:   0.000
 Max.   :19352.00   Max.   :5153.000   Max.   :11739.00   Max.   :5139.000                 Max.   :528.0000                         Max.   :5139.000

 Native_Hawaiian_and_Other_Pacific_Islander_alone  Some_Other_Race_alone   Two_or_More_Races   Hispanic_or_Latino   Total_Male_Pop
 Min.   :  0.0000                                   Min.   :   0.000        Min.   :  0.0000    Min.   :   0.000     Min.   :   0.0
 1st Qu.:  0.0000                                   1st Qu.:   0.000        1st Qu.:  0.0000    1st Qu.:   0.000     1st Qu.:   0.0
 Median :  0.0000                                   Median :   0.000        Median :  0.0000    Median :   0.000     Median :   2.0
 Mean   :  0.0378                                   Mean   :   1.252        Mean   :  0.7133    Mean   :   3.454     Mean   :  13.5
 3rd Qu.:  0.0000                                   3rd Qu.:   0.000        3rd Qu.:  0.0000    3rd Qu.:   0.000     3rd Qu.:  14.0
 Max.   :851.0000                                   Max.   :1452.000        Max.   :721.0000    Max.   :2651.000     Max.   :5144.0

  Male_Under_5     Male_5_to_9     Male_10_to_14     Male_15_to_17    Male_18_and_19     Male_20          Male_21        Male_22_to_24    Male_25_to_29    Male_30_to_34
 Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :   0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :   0.0    Min.   :  0.0
 1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:   0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:   0.0    1st Qu.:  0.0
 Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :   0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :   0.0    Median :  0.0
 Mean   :  0.9    Mean   :  0.9    Mean   :  0.9    Mean   :  0.6    Mean   :   0.4    Mean   :  0.2    Mean   :  0.2    Mean   :  0.6    Mean   :   0.9    Mean   :  0.9
 3rd Qu.:  1.0    3rd Qu.:  0.9    3rd Qu.:  1.0    3rd Qu.:  0.0    3rd Qu.:   0.0    3rd Qu.:  0.0    3rd Qu.:  0.0    3rd Qu.:  0.0    3rd Qu.:   1.0    3rd Qu.:  1.0
 Max.   :2064.0   Max.   :280.0    Max.   :355.0    Max.   :572.0    Max.   :2029.0    Max.   :843.0    Max.   :764.0    Max.   :2651.0   Max.   :5144.0    Max.   :521.0

  Male_35_to_39   Male_40_to_44    Male_45_to_49   Male_50_to_54    Male_55_to_59    Male_60_and_61   Male_62_to_64    Male_65_and_66   Male_67_to_69    Male_70_to_74
 Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0    Min.   :  0.0
 1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0
 Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0    Median :  0.0
 Mean   :  0.9    Mean   :  0.9    Mean   :  0.9    Mean   :  0.9    Mean   :  0.9    Mean   :  0.2    Mean   :  0.4    Mean   :  0.3    Mean   :  0.3    Mean   :  0.9
 3rd Qu.:  0.9    3rd Qu.:  1.0    3rd Qu.:  0.9    3rd Qu.:  1.0    3rd Qu.:  1.0    3rd Qu.:  0.0    3rd Qu.:  0.0    3rd Qu.:  0.3    3rd Qu.:  0.4    3rd Qu.:  1.0
 Max.   :280.0    Max.   :355.0    Max.   :572.0    Max.   :809.0    Max.   :784.0    Max.   :480.0    Max.   :854.0    Max.   :385.0    Max.   :748.0    Max.   :809.0

  Male_75_to_79    Male_80_to_84    Male_85_and_over   Female_Under_5   Total_Female_Pop   Female_5_to_9    Female_10_to_14   Female_15_to_17   Female_18_and_19
 Min.   :  0.0    Min.   :  0.0    Min.   :  0.0      Min.   :  0.0    Min.   :   0.0     Min.   :  0.0    Min.   :  0.0     Min.   :  0.0     Min.   :  0.0
 1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0      1st Qu.:  0.0    1st Qu.:   0.0     1st Qu.:  0.0    1st Qu.:  0.0     1st Qu.:  0.0     1st Qu.:  0.0
 Median :  0.0    Median :  0.0    Median :  0.0      Median :  2.0    Median :   2.0     Median :  0.0    Median :  0.0     Median :  0.0     Median :  0.0
 Mean   :  0.3    Mean   :  0.2    Mean   :  0.6      Mean   : 14.2    Mean   :   0.9     Mean   :  0.9    Mean   :  0.9     Mean   :  0.6     Mean   :  0.4
 3rd Qu.:  0.0    3rd Qu.:  0.0    3rd Qu.:  0.2      3rd Qu.: 15.0    3rd Qu.:   1.0     3rd Qu.:  1.0    3rd Qu.:  0.9     3rd Qu.:  0.6     3rd Qu.:  0.0
 Max.   :175.0    Max.   :199.0    Max.   :355.0      Max.   :2710.0   Max.   :   1.0     Max.   :476.0    Max.   :288.0     Max.   :252.0     Max.   :1902.0

   Female_20        Female_21        Female_22_to_24   Female_25_to_29   Female_30_to_34   Female_35_to_39   Female_40_to_44   Female_45_to_49   Female_50_to_54
 Min.   :  0.0    Min.   :  0.0    Min.   :  0.0     Min.   :  0.0     Min.   :  0.0     Min.   :  0.0     Min.   :  0.0     Min.   :  0.0     Min.   :  0.0
 1st Qu.:  0.0    1st Qu.:  0.0    1st Qu.:  0.0     1st Qu.:  0.0     1st Qu.:  0.0     1st Qu.:  0.0     1st Qu.:  0.0     1st Qu.:  0.0     1st Qu.:  0.0
 Median :  0.0    Median :  0.0    Median :  0.0     Median :  0.0     Median :  0.0     Median :  0.0     Median :  0.0     Median :  0.0     Median :  0.0
 Mean   :  0.2    Mean   :  0.2    Mean   :  0.6     Mean   :  0.9     Mean   :  0.9     Mean   :  0.9     Mean   :  0.9     Mean   :  0.9     Mean   :  1.1
 3rd Qu.:  0.0    3rd Qu.:  0.0    3rd Qu.:  0.6     3rd Qu.:  1.0     3rd Qu.:  1.0     3rd Qu.:  1.0     3rd Qu.:  1.0     3rd Qu.:  1.0     3rd Qu.:  1.0
 Max.   :732.0    Max.   :494.0    Max.   :413.0     Max.   :519.0     Max.   :437.0     Max.   :380.0     Max.   :372.0     Max.   :344.0     Max.   :219.0
```

Figure 4: Cleaned Data Summary pt.1

```
 Female_55_to_59   Female_60_and_61  Female_62_to_64   Female_65_and_66  Female_67_to_69   Female_70_to_74   Female_75_to_79   Female_80_to_84   Female_85_and_over
 Min.   :   0.00   Min.   :   0.0    Min.   :   0.0    Min.   :   0.0    Min.   :   0.0    Min.   :   0.0    Min.   :   0.0    Min.   :   0.0    Min.   :   0.0
 1st Qu.:   0.00   1st Qu.:   0.0    1st Qu.:   0.0    1st Qu.:   0.0    1st Qu.:   0.0    1st Qu.:   0.0    1st Qu.:   0.0    1st Qu.:   0.0    1st Qu.:   0.0
 Median :   2.00   Median :   0.3    Median :   0.5    Median :   0.3    Median :   0.4    Median :   0.5    Median :   0.4    Median :   0.3    Median :   0.4
 Mean   :  10.43   Mean   :   0.9    Mean   :   0.5    Mean   :   0.3    Mean   :   0.4    Mean   :   0.5    Mean   :   0.4    Mean   :   0.3    Mean   :   0.4
 3rd Qu.:  11.00   3rd Qu.:   1.0    3rd Qu.:   0.0    3rd Qu.:   0.0    3rd Qu.:   0.0    3rd Qu.:   0.0    3rd Qu.:   0.0    3rd Qu.:   0.0    3rd Qu.:   0.0
 Max.   :2565.00   Max.   :196.0     Max.   :150.0     Max.   :129.0     Max.   :190.0     Max.   :207.0     Max.   :251.0     Max.   :398.0     Max.   :692.0

 Total_Households   Total_1_Person_Households  Total_2_Person_Households  Total_3_Person_Households  Total_4_Person_Households  Total_5_Person_Households
 Min.   :   0.000   Min.   :   0.000           Min.   :   0.000           Min.   :   0.000           Min.   :   0.000           Min.   :   0.000
 1st Qu.:   0.000   1st Qu.:   0.000           1st Qu.:   0.000           1st Qu.:   0.000           1st Qu.:   0.000           1st Qu.:   0.000
 Median :   2.000   Median :   0.000           Median :   0.000           Median :   0.000           Median :   0.000           Median :   0.000
 Mean   :  10.43    Mean   :   2.891           Mean   :   3.438           Mean   :   1.672           Mean   :   1.354           Mean   :   0.6425
 3rd Qu.:  11.00    3rd Qu.:   2.000           3rd Qu.:   4.000           3rd Qu.:   2.000           3rd Qu.:   1.000           3rd Qu.:   1.0000
 Max.   :2565.00    Max.   :1549.000           Max.   :1003.000           Max.   : 355.000           Max.   : 355.000           Max.   : 134.0000

 Total_6_Person_Households  Total_7_or_More_Person_Households  Total_Education_Pop  Male_Edu_Pop      Male_No_schooling_completed  Male_Nursery_to_4th_grade
 Min.   :   0.0000          Min.   :   0.0000                  Min.   :   0.000     Min.   :   0.000  Min.   :   0.0000            Min.   :   0.0000
 1st Qu.:   0.0000          1st Qu.:   0.0000                  1st Qu.:   0.000     1st Qu.:   0.000  1st Qu.:   0.0000            1st Qu.:   0.0000
 Median :   0.0000          Median :   0.0000                  Median :   2.734     Median :   1.342  Median :   0.0000            Median :   0.0000
 Mean   :   0.2555          Mean   :   0.1778                  Mean   :  19.107     Mean   :   9.195  Mean   :   0.1249            Mean   :   0.0567
 3rd Qu.:   0.0000          3rd Qu.:   0.0000                  3rd Qu.:  19.877     3rd Qu.:   9.548  3rd Qu.:   0.0476            3rd Qu.:   0.0000
 Max.   : 126.0000          Max.   : 162.0000                  Max.   :5985.000     Max.   :5985.000  Max.   : 280.5124            Max.   : 280.5124

 Male_5th_and_6th_grade  Male_7th_and_8th_grade  Male_9th_grade    Male_10th_grade   Male_11th_grade   Male_12th_grade_no_diploma
 Min.   :   0.0000       Min.   :   0.0000       Min.   :   0.0000  Min.   :   0.0000  Min.   :   0.0000  Min.   :   0.0000
 1st Qu.:   0.0000       1st Qu.:   0.0000       1st Qu.:   0.0000  1st Qu.:   0.0000  1st Qu.:   0.0000  1st Qu.:   0.0000
 Median :   0.0000       Median :   0.0000       Median :   0.0000  Median :   0.0000  Median :   0.0000  Median :   0.0000
 Mean   :   0.1534       Mean   :   0.1778       Mean   :   0.1447  Mean   :   0.2317  Mean   :   0.1755  Mean   :   0.1960
 3rd Qu.:   0.0794       3rd Qu.:   0.0794       3rd Qu.:   0.0617  3rd Qu.:   0.0726  3rd Qu.:   0.1090  3rd Qu.:   0.1246
 Max.   : 348.8686       Max.   : 162.0000       Max.   : 350.4022  Max.   : 312.0523  Max.   : 469.3890  Max.   : 765.0000

 Male_High_school_graduate_includes_equivalency  Male_Some_college_less_than_1_year  Male_Some_college_1_or_more_years_no_degree  Male_Associates_degree
 Min.   :   0.0000                               Min.   :   0.0000                   Min.   :   0.0000                            Min.   :   0.0000
 1st Qu.:   0.0000                               1st Qu.:   0.0000                   1st Qu.:   0.0000                            1st Qu.:   0.0000
 Median :   0.0000                               Median :   0.0691                   Median :   0.0819                            Median :   0.0000
 Mean   :   0.4334                               Mean   :   0.5546                   Mean   :   0.6939                            Mean   :   1.2640
 3rd Qu.:   0.4047                               3rd Qu.:   0.5254                   3rd Qu.:   0.6383                            3rd Qu.:   1.2152
 Max.   : 536.7497                               Max.   : 522.0000                   Max.   : 406.4220                            Max.   : 907.8602

 Male_Bachelors_degree  Male_Masters_degree  Male_Professional_school_degree  Male_Doctorate_degree  Female_Edu_Pop    Female_No_schooling_completed
 Min.   :   0.0000      Min.   :   0.0000    Min.   :   0.0000                Min.   :   0.0000      Min.   :   0.000  Min.   :   0.0000
 1st Qu.:   0.1329      1st Qu.:   0.0290    1st Qu.:   0.0000                1st Qu.:   0.0000      1st Qu.:   0.000  1st Qu.:   0.0000
 Median :   0.7299      Median :   0.7106    Median :   0.0000                Median :   0.0000      Median :   1.372  Median :   0.0000
 Mean   :   1.7299      Mean   :   0.7106    Mean   :   0.2317                Mean   :   0.1531      Mean   :   9.912  Mean   :   0.1303
 3rd Qu.:   1.2120      3rd Qu.:   0.3874    3rd Qu.:   0.0726                3rd Qu.:   0.0382      3rd Qu.:  10.255  3rd Qu.:   0.0436
 Max.   : 774.7331      Max.   : 525.8827    Max.   : 312.0523                Max.   : 218.1944      Max.   :3253.455  Max.   : 141.2620

 Female_Nursery_to_4th_grade  Female_5th_and_6th_grade  Female_7th_and_8th_grade  Female_9th_grade  Female_10th_grade  Female_11th_grade  Female_12th_grade_no_diploma
 Min.   :   0.0000            Min.   :   0.0000         Min.   :   0.0000         Min.   :   0.0000  Min.   :   0.0000  Min.   :   0.0000  Min.   :   0.0000
 1st Qu.:   0.0000            1st Qu.:   0.0000         1st Qu.:   0.0000         1st Qu.:   0.0000  1st Qu.:   0.0000  1st Qu.:   0.0000  1st Qu.:   0.0000
 Median :   0.0000            Median :   0.0000         Median :   0.0000         Median :   0.0000  Median :   0.0000  Median :   0.0000  Median :   0.0000
 Mean   :   0.0579            Mean   :   0.1103         Mean   :   0.1497         Mean   :   0.1316  Mean   :   0.1657  Mean   :   0.1871  Mean   :   0.1552
 3rd Qu.:   0.0000            3rd Qu.:   0.0000         3rd Qu.:   0.0741         3rd Qu.:   0.0505  3rd Qu.:   0.1021  3rd Qu.:   0.1160  3rd Qu.:   0.0906
 Max.   : 105.0000            Max.   : 158.0659         Max.   : 134.2717         Max.   : 299.4785  Max.   : 190.8764  Max.   : 270.5179  Max.   : 157.9311
```

Figure 5: Cleaned Data Summary pt.2

```
 Female_High_school_graduate_includes_equivalency Female_Some_college_less_than_1_year Female_Some_college_1_or_more_years_no_degree Female_Associates_degree
 Min.   :   0.0000                                Min.   :  0.0000                     Min.   :  0.0000                              Min.   :  0.0000
 1st Qu.:   0.0000                                1st Qu.:  0.0000                     1st Qu.:  0.0000                              1st Qu.:  0.0000
 Median :   0.4066                                Median :  0.0808                     Median :  0.1803                              Median :  0.1208
 Mean   :   2.7196                                Mean   :  0.6335                     Mean   :  1.3676                              Mean   :  0.9261
 3rd Qu.:   2.8136                                3rd Qu.:  0.6123                     3rd Qu.:  1.3255                              3rd Qu.:  0.8983
 Max.   :1151.1820                                Max.   :211.2233                     Max.   :722.0393                              Max.   :260.0000

 Female_Bachelors_degree  Female_Masters_degree  Female_Professional_school_degree  Female_Doctorate_degree  Median_household_income_in_2019_dollars
 Min.   :  0.0000         Min.   :  0.0000       Min.   :  0.0000                   Min.   :  0.0000         Min.   :-666666666
 1st Qu.:  0.0000         1st Qu.:  0.0000       1st Qu.:  0.0000                   1st Qu.:  0.0000         1st Qu.:     44313
 Median :  0.1666         Median :  0.0570       Median :  0.0000                   Median :  0.0000         Median :     55598
 Mean   :  1.9313         Mean   :  0.9578       Mean   :  0.1741                   Mean   :  0.1146         Mean   :   -618687
 3rd Qu.:  1.4537         3rd Qu.:  0.6066       3rd Qu.:  0.0505                   3rd Qu.:  0.0214         3rd Qu.:     69920
 Max.   :888.5101         Max.   :688.5627       Max.   :314.9206                   Max.   :157.8953         Max.   :    250001

 Households_with_no_Internet_access Total_Employment_Pop In_labor_force  In_labor_force_Civilian_labor_force In_labor_force_Civilian_labor_force_Employed
 Min.   :   0.000                   Min.   :   0.000     Min.   :   0.000 Min.   :   0.000                    Min.   :   0.000
 1st Qu.:   0.2495                  1st Qu.:   0.000     1st Qu.:   0.000 1st Qu.:   0.000                    1st Qu.:   0.000
 Median :   3.158                   Median :   1.838     Median :   1.733 Median :   1.733                    Median :   1.733
 Mean   :  22.464                   Mean   :  22.464     Mean   :  14.183 Mean   :  14.183                    Mean   :  13.426
 3rd Qu.:  23.259                   3rd Qu.:  14.209     3rd Qu.:  14.155 3rd Qu.:  14.155                    3rd Qu.:  13.341
 Max.   :6898.426                   Max.   :4845.859     Max.   :4493.750 Max.   :4493.750                    Max.   :4232.317

 In_labor_force_civilian_labor_force_Unemployed In_labor_force_Armed_Forces Not_in_labor_force Public_transportation_excluding_taxicab Taxicab          Motorcycle
 Min.   :   0.0000                              Min.   :   0.000            Min.   :   0.000   Min.   :   0.0000                       Min.   :   0.0000 Min.   :  0.000
 1st Qu.:   0.0000                              1st Qu.:   0.000            1st Qu.:   0.000   1st Qu.:   0.0000                       1st Qu.:   0.0000 1st Qu.:  0.000
 Median :   0.0646                              Median :   0.000            Median :   1.222   Median :   0.0000                       Median :   0.0000 Median :  0.000
 Mean   :   0.7566                              Mean   :   0.075            Mean   :   8.206   Mean   :   0.8091                       Mean   :   0.0277 Mean   :  0.017
 3rd Qu.:   0.6250                              3rd Qu.:   0.000            3rd Qu.:   8.612   3rd Qu.:   0.0337                       3rd Qu.:   0.0000 3rd Qu.:  0.000
 Max.   : 450.3072                              Max.   :4400.859            Max.   :6856.000   Max.   :1669.4557                       Max.   : 193.3018 Max.   : 52.197

 Bicycle           walked            Other_means       Worked_from_home  Racers    Racers_Male  Racers_Female Racers_Junior
 Min.   :   0.0000 Min.   :   0.0000 Min.   :   0.0000 Min.   :   0.0000 Min.   : 0 Min.   : 0   Min.   : 0    Min.   : 0
 1st Qu.:   0.0000 1st Qu.:   0.0000 1st Qu.:   0.0000 1st Qu.:   0.0000 1st Qu.: 0 1st Qu.: 0   1st Qu.: 0    1st Qu.: 0
 Median :   0.0000 Median :   0.0000 Median :   0.0000 Median :   0.0455 Median : 0 Median : 1   Median : 0    Median : 0
 Mean   :   0.0709 Mean   :   0.3825 Mean   :   0.1167 Mean   :   0.6303 Mean   : 0 Mean   : 1   Mean   : 0    Mean   : 0
 3rd Qu.:   0.0000 3rd Qu.:   0.1471 3rd Qu.:   0.0398 3rd Qu.:   0.4574 3rd Qu.: 0 3rd Qu.: 1   3rd Qu.: 0    3rd Qu.: 0
 Max.   : 138.9354 Max.   :2181.3190 Max.   : 114.3129 Max.   :1604.5976 Max.   :17 Max.   :39   Max.   :13    Max.   :22

 Racers_Master Racers_Pro_Road Racers_Pro_MTB Racers_Pro_Track Racers_Pro_Cross Racers_Mid_road Racers_Mid_MTB Racers_Mid_Track Racers_Mid_Cross Racers_Elite
 Min.   : 0    Min.   :0       Min.   : 0     Min.   : 0       Min.   : 0       Min.   : 0      Min.   : 0     Min.   : 0       Min.   : 0       Min.   : 0
 1st Qu.: 0    1st Qu.:0       1st Qu.: 0     1st Qu.: 0       1st Qu.: 0       1st Qu.: 0      1st Qu.: 0     1st Qu.: 0       1st Qu.: 0       1st Qu.: 0
 Median : 0    Median :0       Median : 0     Median : 0       Median : 0       Median : 0      Median : 1     Median : 0       Median : 0       Median : 0
 Mean   : 1    Mean   :0       Mean   : 0     Mean   : 0       Mean   : 0       Mean   : 0      Mean   : 1     Mean   : 0       Mean   : 0       Mean   : 0
 3rd Qu.: 1    3rd Qu.:0       3rd Qu.: 0     3rd Qu.: 0       3rd Qu.: 0       3rd Qu.: 0      3rd Qu.: 1     3rd Qu.: 0       3rd Qu.: 0       3rd Qu.: 0
 Max.   :25    Max.   :9       Max.   :27     Max.   : 4       Max.   :13       Max.   :22      Max.   :48     Max.   :13       Max.   :26       Max.   :15
```

Figure 6: Cleaned Data Summary pt.3

```
 Male_Racers      Female_Racers    Junior_Racers     Elite_Racers      Master_Racers     Pro_Road_Racers   Pro_MTB_Racers    Pro_Track_Racers  Pro_Cross_Racers
 Min.   : 0       Min.   :0        Min.   :0         Min.   :0         Min.   : 0        Min.   :0         Min.   :0         Min.   :0         Min.   :0
 1st Qu.: 0       1st Qu.:0        1st Qu.:0         1st Qu.:0         1st Qu.: 0        1st Qu.:0         1st Qu.:0         1st Qu.:0         1st Qu.:0
 Median : 0       Median :0        Median :0         Median :0         Median : 0        Median :0         Median :0         Median :0         Median :0
 Mean   : 0       Mean   :0        Mean   :0         Mean   :0         Mean   : 0        Mean   :0         Mean   :0         Mean   :0         Mean   :0
 3rd Qu.: 0       3rd Qu.:0        3rd Qu.:0         3rd Qu.:0         3rd Qu.: 0        3rd Qu.:0         3rd Qu.:0         3rd Qu.:0         3rd Qu.:0
 Max.   :12       Max.   :5        Max.   :3         Max.   :4         Max.   :12        Max.   :2         Max.   :4         Max.   :1         Max.   :3
 Mid_Road_Racers  Mid_MTB_Racers   Mid_Track_Racers  Mid_Cross_Racers     Min_Temp
 Min.   :0        Min.   : 0       Min.   :0         Min.   :0        Min.   :-0.30
 1st Qu.:0        1st Qu.: 0       1st Qu.:0         1st Qu.:0        1st Qu.:19.70
 Median :0        Median : 0       Median :0         Median :0        Median :25.40
 Mean   :0        Mean   : 0       Mean   :0         Mean   :0        Mean   :27.14
 3rd Qu.:0        3rd Qu.: 0       3rd Qu.:0         3rd Qu.:0        3rd Qu.:33.40
 Max.   :2        Max.   :17       Max.   :1         Max.   :3        Max.   :61.30
```

Figure 7: Cleaned Data Summary pt.4