test

2019 3 13

1. ggplot2 midwest .

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
df_midwest = midwest

str(df_midwest)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    437 obs. of  28 variables:
##  $ PID                 : int  561 562 563 564 565 566 567 568 569 570 ...
##  $ county              : chr  "ADAMS" "ALEXANDER" "BOND" "BOONE" ...
##  $ state               : chr  "IL" "IL" "IL" "IL" ...
##  $ area                : num  0.052 0.014 0.022 0.017 0.018 0.05 0.017 0.027 0.024 0.058 ...
##  $ poptotal            : int  66090 10626 14991 30806 5836 35688 5322 16805 13437 173025 ...
##  $ popdensity          : num  1271 759 681 1812 324 ...
##  $ popwhite            : int  63917 7054 14477 29344 5264 35157 5298 16519 13384 146506 ...
##  $ popblack            : int  1702 3496 429 127 547 50 1 111 16 16559 ...
##  $ popamerindian       : int  98 19 35 46 14 65 8 30 8 331 ...
##  $ popasian            : int  249 48 16 150 5 195 15 61 23 8033 ...
##  $ popother            : int  124 9 34 1139 6 221 0 84 6 1596 ...
##  $ percwhite           : num  96.7 66.4 96.6 95.3 90.2 ...
##  $ percblack           : num  2.575 32.9 2.862 0.412 9.373 ...
##  $ percamerindan       : num  0.148 0.179 0.233 0.149 0.24 ...
##  $ percasian           : num  0.3768 0.4517 0.1067 0.4869 0.0857 ...
##  $ percother           : num  0.1876 0.0847 0.2268 3.6973 0.1028 ...
##  $ popadults           : int  43298 6724 9669 19272 3979 23444 3583 11323 8825 95971 ...
##  $ perchsd             : num  75.1 59.7 69.3 75.5 68.9 ...
##  $ percollege          : num  19.6 11.2 17 17.3 14.5 ...
##  $ percprof            : num  4.36 2.87 4.49 4.2 3.37 ...
##  $ poppovertyknown     : int  63628 10529 14235 30337 4815 35107 5241 16455 13081 154934 ...
##  $ percpovertyknown    : num  96.3 99.1 95 98.5 82.5 ...
##  $ percbelowpoverty    : num  13.15 32.24 12.07 7.21 13.52 ...
##  $ percchildbelowpovert: num  18 45.8 14 11.2 13 ...
```

```
##  $ percadultpoverty    : num  11.01 27.39 10.85 5.54 11.14 ...
##  $ percelderlypoverty  : num  12.44 25.23 12.7 6.22 19.2 ...
##  $ inmetro             : int  0 0 0 1 0 0 0 0 0 1 ...
##  $ category            : chr  "AAR" "LHR" "AAR" "ALU" ...
```

```
head(df_midwest)
```

```
## # A tibble: 6 x 28
##     PID county state  area poptotal popdensity popwhite popblack
##   <int> <chr>  <chr> <dbl>    <int>      <dbl>    <int>    <int>
## 1   561 ADAMS  IL    0.052    66090      1271.    63917     1702
## 2   562 ALEXA~ IL    0.014    10626       759      7054     3496
## 3   563 BOND   IL    0.022    14991       681.    14477      429
## 4   564 BOONE  IL    0.017    30806      1812.    29344      127
## 5   565 BROWN  IL    0.018     5836       324.     5264      547
## 6   566 BUREAU IL    0.05     35688       714.    35157       50
## # ... with 20 more variables: popamerindian <int>, popasian <int>,
## #   popother <int>, percwhite <dbl>, percblack <dbl>, percamerindan <dbl>,
## #   percasian <dbl>, percother <dbl>, popadults <int>, perchsd <dbl>,
## #   percollege <dbl>, percprof <dbl>, poppovertyknown <int>,
## #   percpovertyknown <dbl>, percbelowpoverty <dbl>,
## #   percchildbelowpovert <dbl>, percadultpoverty <dbl>,
## #   percelderlypoverty <dbl>, inmetro <int>, category <chr>
```

```
tail(df_midwest)
```

```
## # A tibble: 6 x 28
##     PID county state  area poptotal popdensity popwhite popblack
##   <int> <chr>  <chr> <dbl>    <int>      <dbl>    <int>    <int>
## 1  3047 WASHI~ WI    0.025    95328      3813.    94465      125
## 2  3048 WAUKE~ WI    0.034   304715      8962.   298313     1096
## 3  3049 WAUPA~ WI    0.045    46104      1025.    45695       22
## 4  3050 WAUSH~ WI    0.037    19385       524.    19094       29
## 5  3051 WINNE~ WI    0.035   140320      4009.   136822      697
## 6  3052 WOOD   WI    0.048    73605      1533.    72157       90
## # ... with 20 more variables: popamerindian <int>, popasian <int>,
## #   popother <int>, percwhite <dbl>, percblack <dbl>, percamerindan <dbl>,
## #   percasian <dbl>, percother <dbl>, popadults <int>, perchsd <dbl>,
## #   percollege <dbl>, percprof <dbl>, poppovertyknown <int>,
## #   percpovertyknown <dbl>, percbelowpoverty <dbl>,
## #   percchildbelowpovert <dbl>, percadultpoverty <dbl>,
## #   percelderlypoverty <dbl>, inmetro <int>, category <chr>
```

```
summary(df_midwest)
```

```
##       PID          county             state                area         
##  Min.   : 561   Length:437         Length:437         Min.   :0.00500  
##  1st Qu.: 670   Class :character   Class :character   1st Qu.:0.02400  
##  Median :1221   Mode  :character   Mode  :character   Median :0.03000  
##  Mean   :1437                                         Mean   :0.03317  
##  3rd Qu.:2059                                         3rd Qu.:0.03800  
##  Max.   :3052                                         Max.   :0.11000
```

```
##     poptotal         popdensity          popwhite          popblack
##  Min.   :   1701   Min.   :   85.05   Min.   :    416   Min.   :      0
##  1st Qu.:  18840   1st Qu.:  622.41   1st Qu.:  18630   1st Qu.:     29
##  Median :  35324   Median : 1156.21   Median :  34471   Median :    201
##  Mean   :  96130   Mean   : 3097.74   Mean   :  81840   Mean   :  11024
##  3rd Qu.:  75651   3rd Qu.: 2330.00   3rd Qu.:  72968   3rd Qu.:   1291
##  Max.   :5105067   Max.   :88018.40   Max.   :3204947   Max.   :1317147
##  popamerindian       popasian         popother        percwhite
##  Min.   :    4.0   Min.   :     0   Min.   :     0   Min.   :10.69
##  1st Qu.:   44.0   1st Qu.:    35   1st Qu.:    20   1st Qu.:94.89
##  Median :   94.0   Median :   102   Median :    66   Median :98.03
##  Mean   :  343.1   Mean   :  1310   Mean   :  1613   Mean   :95.56
##  3rd Qu.:  288.0   3rd Qu.:   401   3rd Qu.:   345   3rd Qu.:99.07
##  Max.   :10289.0   Max.   :188565   Max.   :384119   Max.   :99.82
##    percblack       percamerindan        percasian        percother
##  Min.   : 0.0000   Min.   : 0.05623   Min.   :0.0000   Min.   :0.00000
##  1st Qu.: 0.1157   1st Qu.: 0.15793   1st Qu.:0.1737   1st Qu.:0.09102
##  Median : 0.5390   Median : 0.21502   Median :0.2972   Median :0.17844
##  Mean   : 2.6763   Mean   : 0.79894   Mean   :0.4872   Mean   :0.47906
##  3rd Qu.: 2.6014   3rd Qu.: 0.38362   3rd Qu.:0.5212   3rd Qu.:0.48050
##  Max.   :40.2100   Max.   :89.17738   Max.   :5.0705   Max.   :7.52427
##    popadults          perchsd        percollege        percprof
##  Min.   :   1287   Min.   :46.91   Min.   : 7.336   Min.   : 0.5203
##  1st Qu.:  12271   1st Qu.:71.33   1st Qu.:14.114   1st Qu.: 2.9980
##  Median :  22188   Median :74.25   Median :16.798   Median : 3.8142
##  Mean   :  60973   Mean   :73.97   Mean   :18.273   Mean   : 4.4473
##  3rd Qu.:  47541   3rd Qu.:77.20   3rd Qu.:20.550   3rd Qu.: 4.9493
##  Max.   :3291995   Max.   :88.90   Max.   :48.079   Max.   :20.7913
##  poppovertyknown   percpovertyknown percbelowpoverty percchildbelowpovert
##  Min.   :   1696   Min.   :80.90    Min.   : 2.180   Min.   : 1.919
##  1st Qu.:  18364   1st Qu.:96.89    1st Qu.: 9.199   1st Qu.:11.624
##  Median :  33788   Median :98.17    Median :11.822   Median :15.270
##  Mean   :  93642   Mean   :97.11    Mean   :12.511   Mean   :16.447
##  3rd Qu.:  72840   3rd Qu.:98.60    3rd Qu.:15.133   3rd Qu.:20.352
##  Max.   :5023523   Max.   :99.86    Max.   :48.691   Max.   :64.308
##  percadultpoverty percelderlypoverty    inmetro         category
##  Min.   : 1.938   Min.   : 3.547     Min.   :0.0000   Length:437
##  1st Qu.: 7.668   1st Qu.: 8.912     1st Qu.:0.0000   Class :character
##  Median :10.008   Median :10.869     Median :0.0000   Mode  :character
##  Mean   :10.919   Mean   :11.389     Mean   :0.3432
##  3rd Qu.:13.182   3rd Qu.:13.412     3rd Qu.:1.0000
##  Max.   :43.312   Max.   :31.162     Max.   :1.0000
```

2. poptotal( ) total , popasian( ) asian .

```
rename(df_midwest, total=poptotal, asian=popasian) -> df_midwest2

names(df_midwest2[5])
```

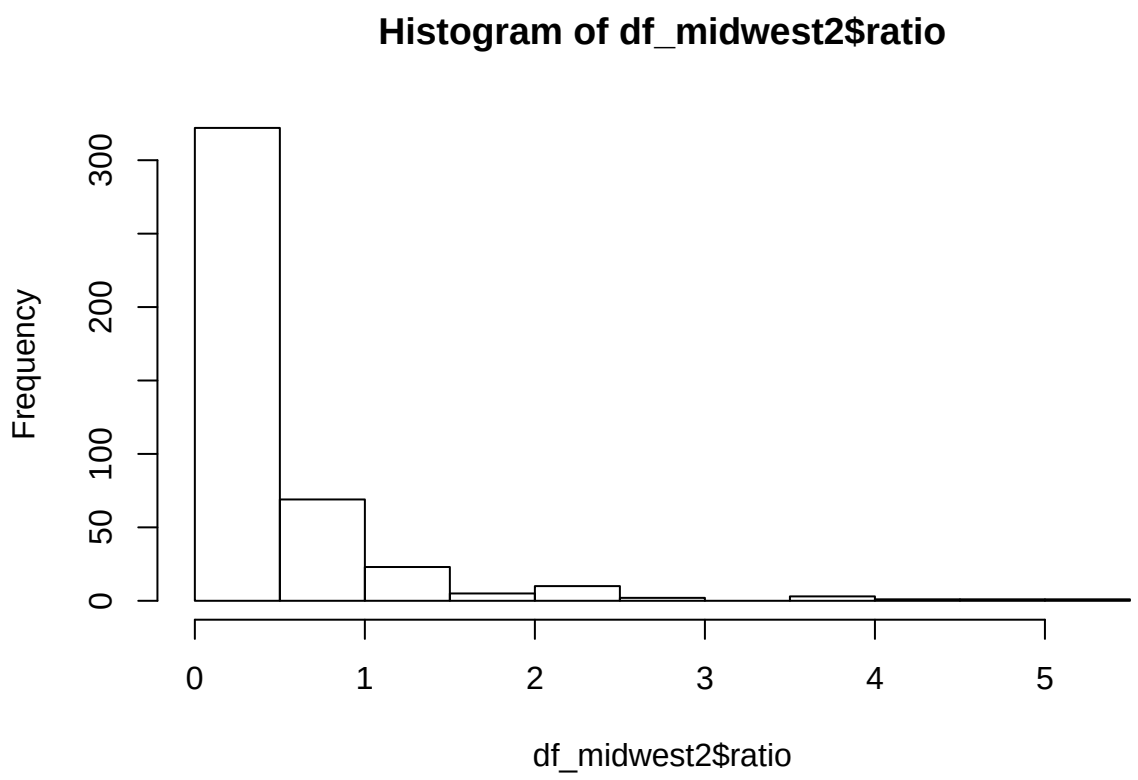
```
## [1] "total"
```

```
names(df_midwest2[10])
```

```
## [1] "asian"
```

3. total, asian ‘ ’ , .

```
df_midwest2$ratio = df_midwest2$asian/df_midwest2$total*100
hist(df_midwest2$ratio)
```



4. , “large”, “small”

```
mean(df_midwest2$ratio)
```

```
## [1] 0.4872462
```

```
df_midwest2$group = ifelse(df_midwest2$ratio > mean(df_midwest2$ratio), "large", "small")
```

5. “lage” “small” , .

```
table(df_midwest2$group)
```

```
##
## large small
##   119   318
```

```
qplot(df_midwest2$group)
```