PLSC 473: American Judicial Behavior

Christopher Zorn

September 22, 2015

Summarizing Data (Univariate Plots / Statistics)

- Numbers
    - Frequency Tables
    - Summary Statistics
- Pictures
    - Histograms
    - Density plots
    - Boxplots

Data: 43 African Countries, 2001

```
> head(Africa)

  ccode cabbr           country population popthou   popden
1   404   GNB     Guinea-Bissau    1227000    1227 0.043471
2   411   GNQ Equatorial Guinea     470000     470 0.019083
3   420   GMB       Gambia, The    1337000    1337 0.133519
4   432   MLI              Mali   11700000   11677 0.009346
5   433   SEN           Senegal    9662000    9662 0.045727
6   434   BEN             Benin    6446000    6446 0.055547
  polity gdppppd tradegdp war adrate healthexp  subsaharan
1      5    0.90    11.25   0    2.8       5.9 Sub-Saharan
2     -5    2.10   272.69   0    3.4       2.0 Sub-Saharan
3     -5    1.77    51.06   0    1.6       6.4 Sub-Saharan
4      6    0.84    12.77   0    1.7       4.3 Sub-Saharan
5      8    1.58    14.20   0    0.5       4.8 Sub-Saharan
6      6    1.04     6.96   0    3.6       4.4 Sub-Saharan
  muslperc literacy internalwar intensity
1       45       42           0         0
2        0       75           1         1
3       90       37           0         0
4       90       28           0         0
5       84       38           1         2
6       20       39           0         0
```

Frequency Tables

- Reports the number of observations that have each value on a given variable.
- Best for *discrete* / *categorical* variables
- Can show frequencies or percentages.

Frequency Tables: Examples

```
> xtabs(~Africa$internalwar)
Africa$internalwar
 0  1
30 13

> # percentages:

> wartab<-xtabs(~Africa$internalwar)
> wartab
Africa$internalwar
 0  1
30 13

> prop.table(wartab)
Africa$internalwar
     0      1
0.6977 0.3023
```

Frequency Tables: Examples

```
# Ordinal variables:

> xtabs(~Africa$intensity)
Africa$intensity
 0  1  2  3
30  6  2  5

> prop.table(xtabs(~Africa$intensity))
Africa$intensity
      0       1       2       3
0.69767 0.13953 0.04651 0.11628



> xtabs(~Africa$polity)
Africa$polity
-9 -7 -6 -5 -4 -3 -2  0  1  2  4  5  6  7  8  9 10
 1  3  4  3  3  1  3  4  2  2  3  3  6  1  1  2  1
```

Summary Statistics

Mean:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Median:

$$M = \left(\frac{N+1}{2}\right) \text{ th term}$$

Standard deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N-1}}$$

Summary Statistics: Examples

```
> mean(Africa$population)
[1] 17388558

> median(Africa$population)
[1] 9662000

> sd(Africa$population)
[1] 22874215
```

```
> mean(Africa$adrate)
[1] 9.365

> median(Africa$adrate)
[1] 6

> sd(Africa$adrate)
[1] 9.96
```

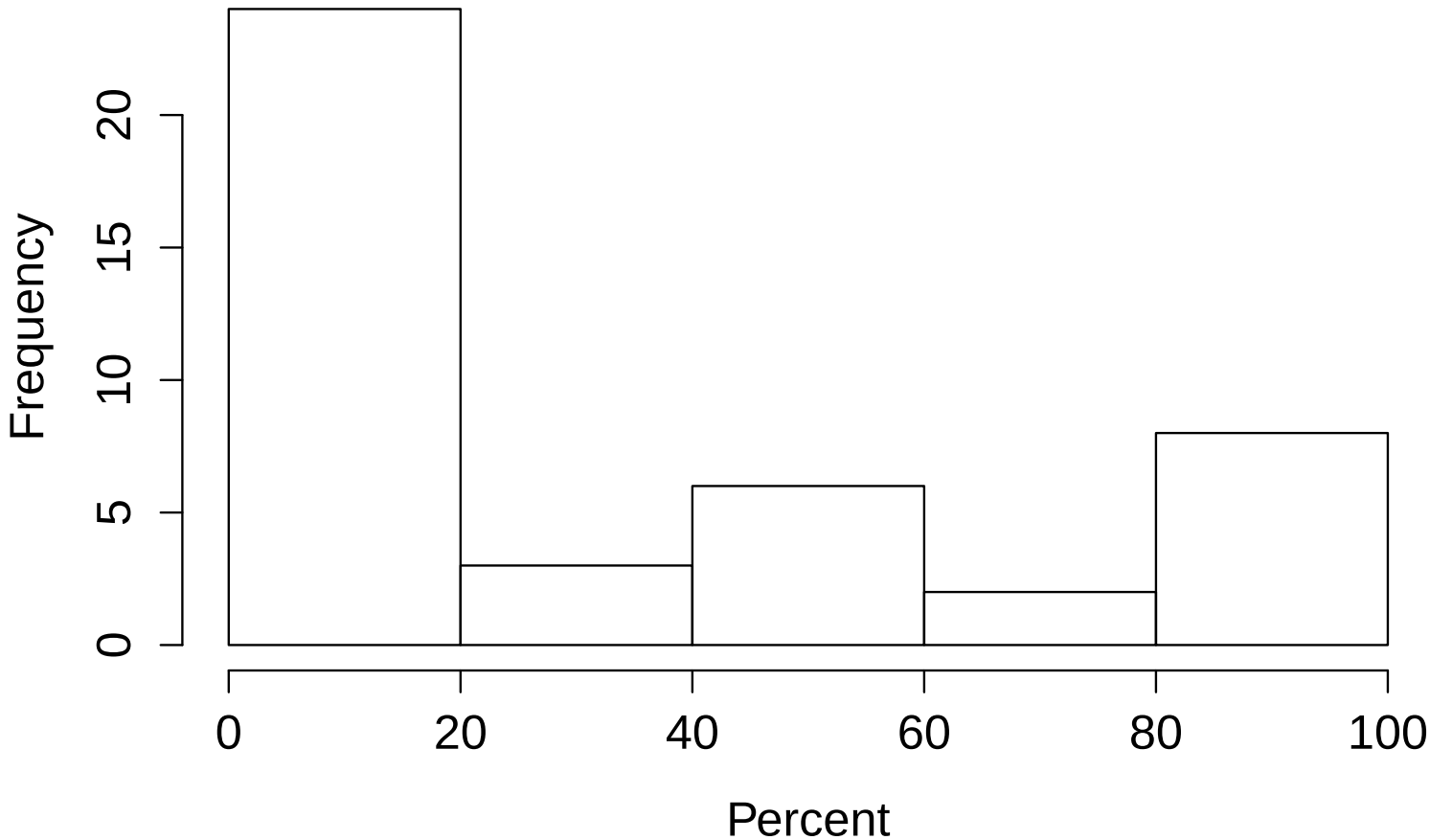
Summary Statistics: Examples

```
> summary(Africa)

     ccode          cabbr                          country     population
 Min.   :404   AGO    : 1   Angola                    : 1   Min.   :   470000
 1st Qu.:452   BDI    : 1   Benin                     : 1   1st Qu.:  3446000
 Median :510   BEN    : 1   Botswana                  : 1   Median :  9662000
 Mean   :510   BWA    : 1   Burundi                   : 1   Mean   : 17388558
 3rd Qu.:556   CAF    : 1   Cameroon                  : 1   3rd Qu.: 19150000
 Max.   :651   CIV    : 1   Central African Republic: 1   Max.   :117000000
               (Other):37   (Other)                   :37
    popthou           popden           polity          gdppppd          tradegdp
 Min.   :   470   Min.   :0.0022   Min.   :-9.000   Min.   : 0.500   Min.   :  4.03
 1st Qu.:  3446   1st Qu.:0.0134   1st Qu.:-4.500   1st Qu.: 0.855   1st Qu.:  7.64
 Median :  9662   Median :0.0357   Median : 0.000   Median : 1.200   Median : 13.56
 Mean   : 17390   Mean   :0.0643   Mean   : 0.512   Mean   : 2.159   Mean   : 30.49
 3rd Qu.: 19189   3rd Qu.:0.0683   3rd Qu.: 5.500   3rd Qu.: 2.040   3rd Qu.: 30.01
 Max.   :116929   Max.   :0.5740   Max.   :10.000   Max.   :10.800   Max.   :272.69

      war             adrate         healthexp              subsaharan    muslperc
 Min.   :0.000   Min.   : 0.10   Min.   :2.00   Not Sub-Saharan: 6   Min.   :  0.0
 1st Qu.:0.000   1st Qu.: 2.70   1st Qu.:3.45   Sub-Saharan    :37   1st Qu.: 10.0
 Median :0.000   Median : 6.00   Median :4.40                        Median : 20.0
 Mean   :0.116   Mean   : 9.37   Mean   :4.60                        Mean   : 36.0
 3rd Qu.:0.000   3rd Qu.:12.90   3rd Qu.:5.80                        3rd Qu.: 55.5
 Max.   :1.000   Max.   :38.80   Max.   :8.60                        Max.   :100.0

    literacy     internalwar      intensity
 Min.   :17.0   Min.   :0.000   Min.   :0.000
 1st Qu.:43.0   1st Qu.:0.000   1st Qu.:0.000
 Median :61.0   Median :0.000   Median :0.000
 Mean   :60.1   Mean   :0.302   Mean   :0.581
 3rd Qu.:78.5   3rd Qu.:1.000   3rd Qu.:1.000
 Max.   :89.0   Max.   :1.000   Max.   :3.000
```

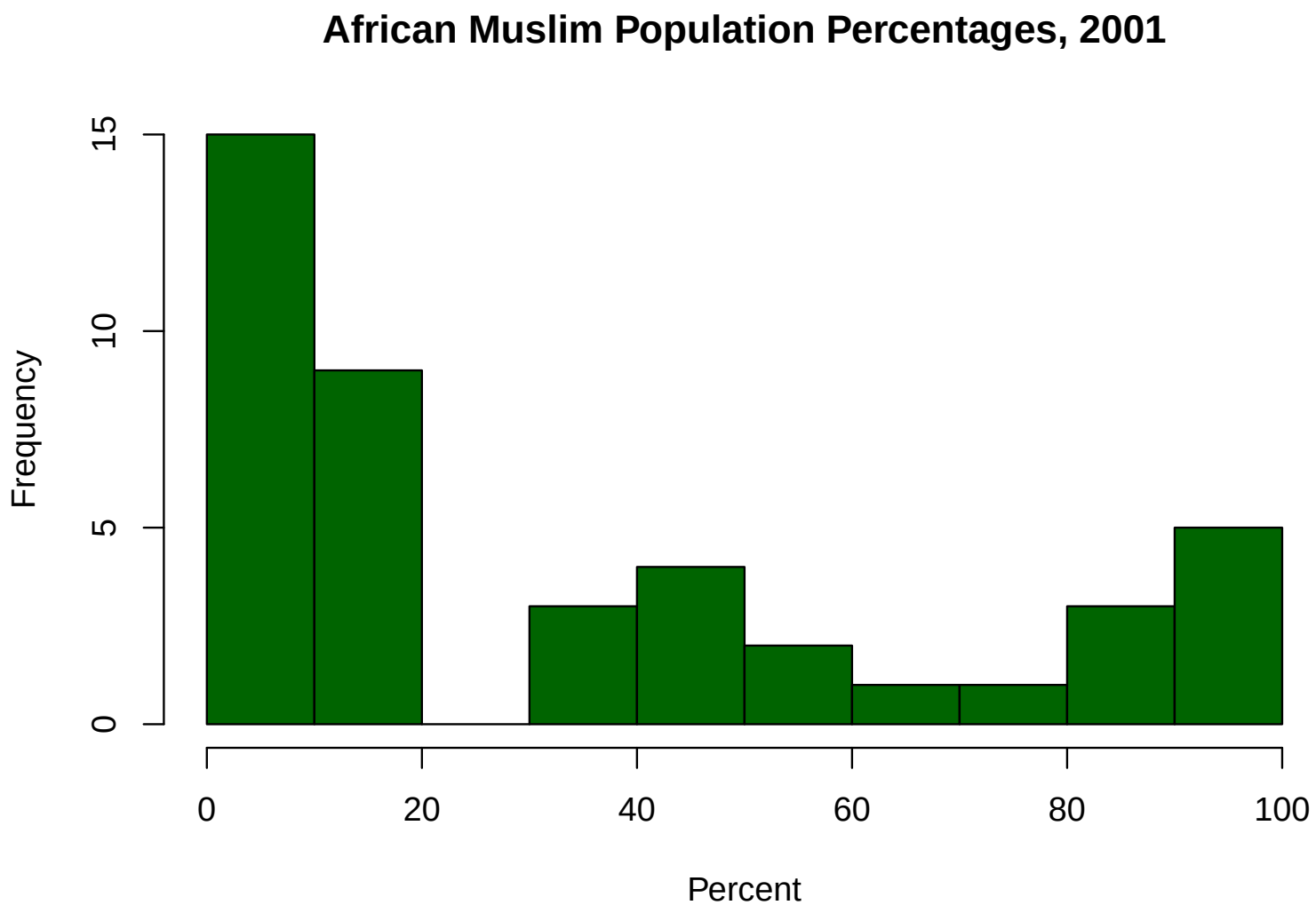
Plots: Histograms

- A "graphical frequency table"
- Can be used with either continuous or discrete variables
- Level of detail set by number of "bins"

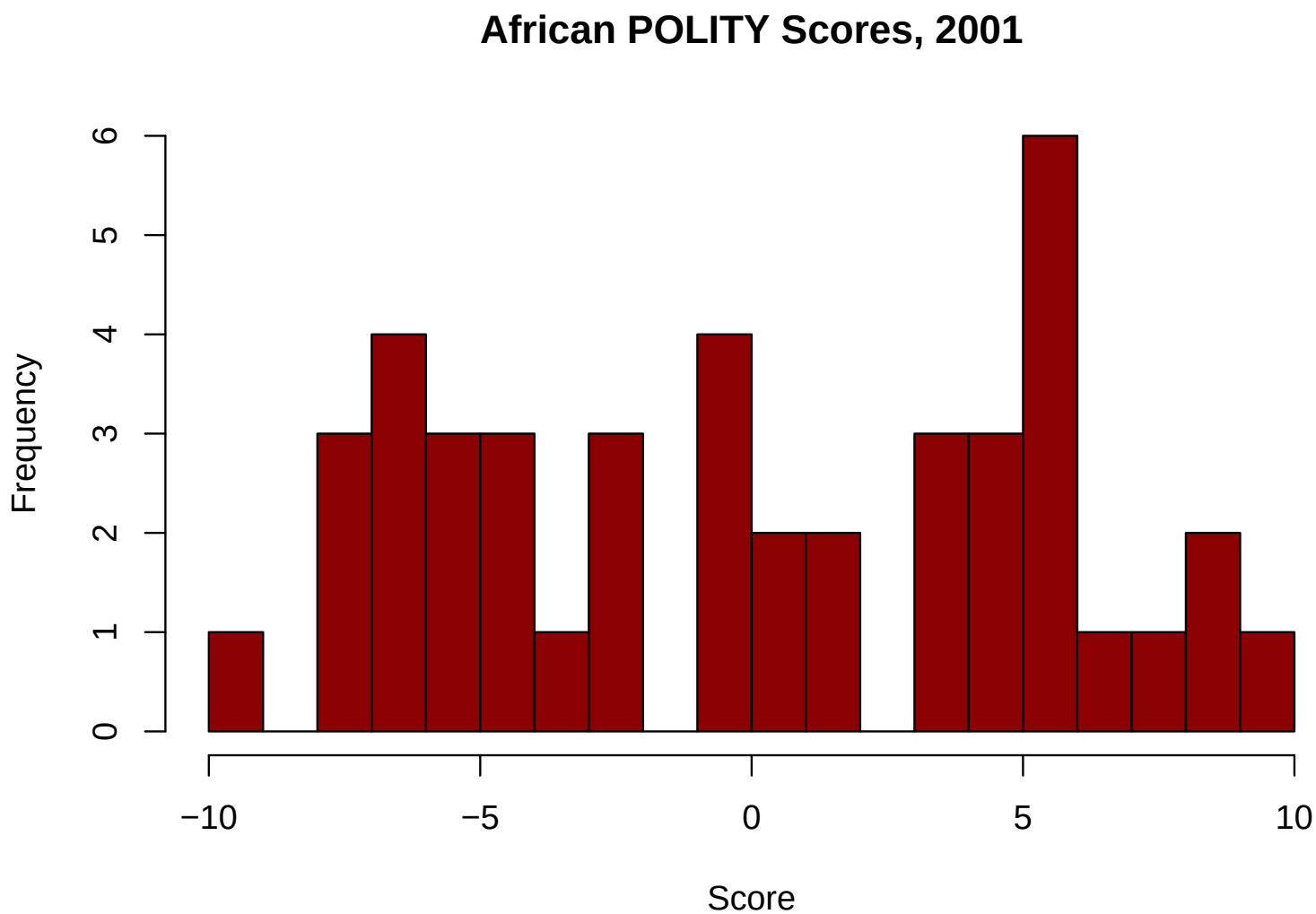
A Basic Histogram



A Much Nicer Histogram



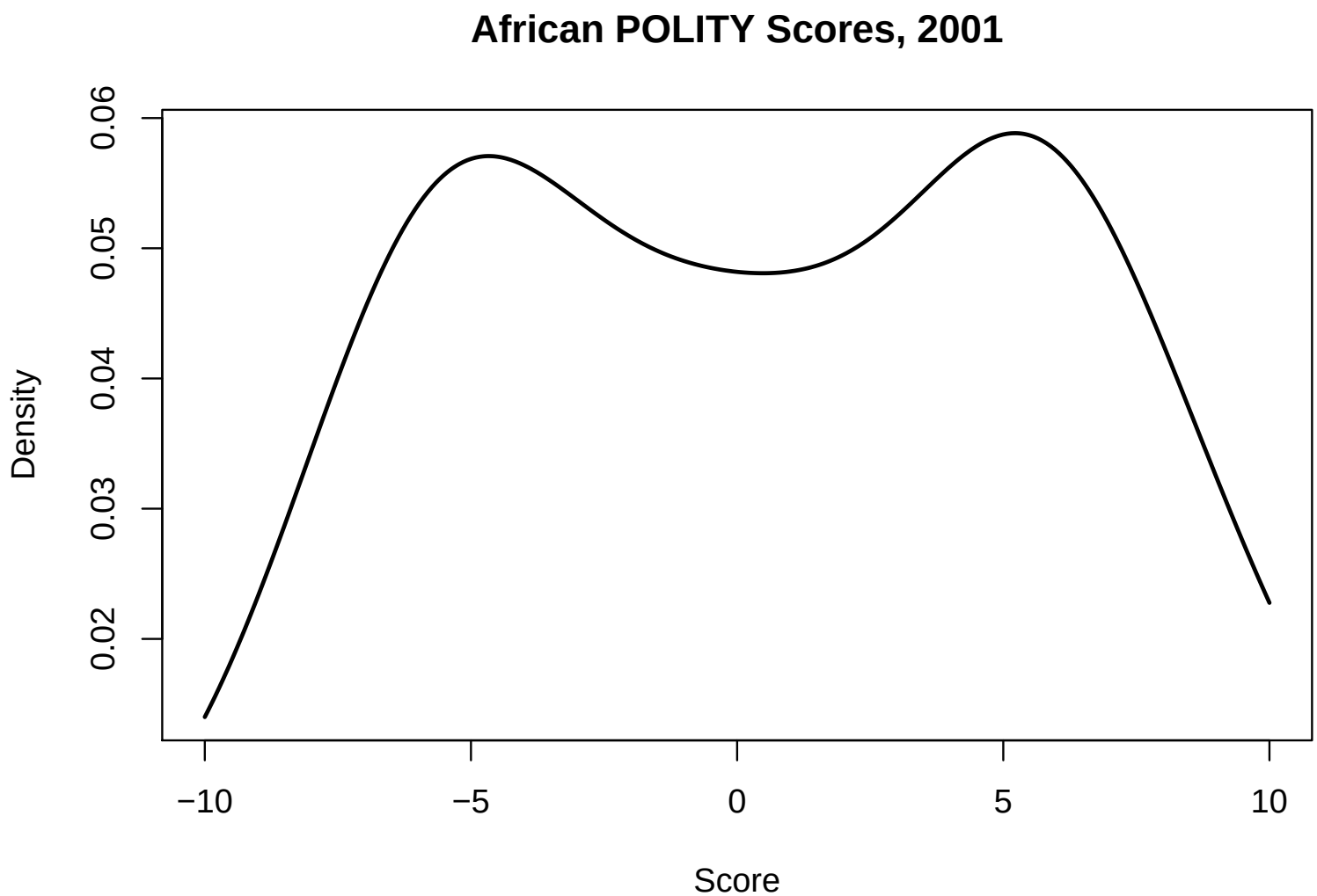
Histogram: POLITY score



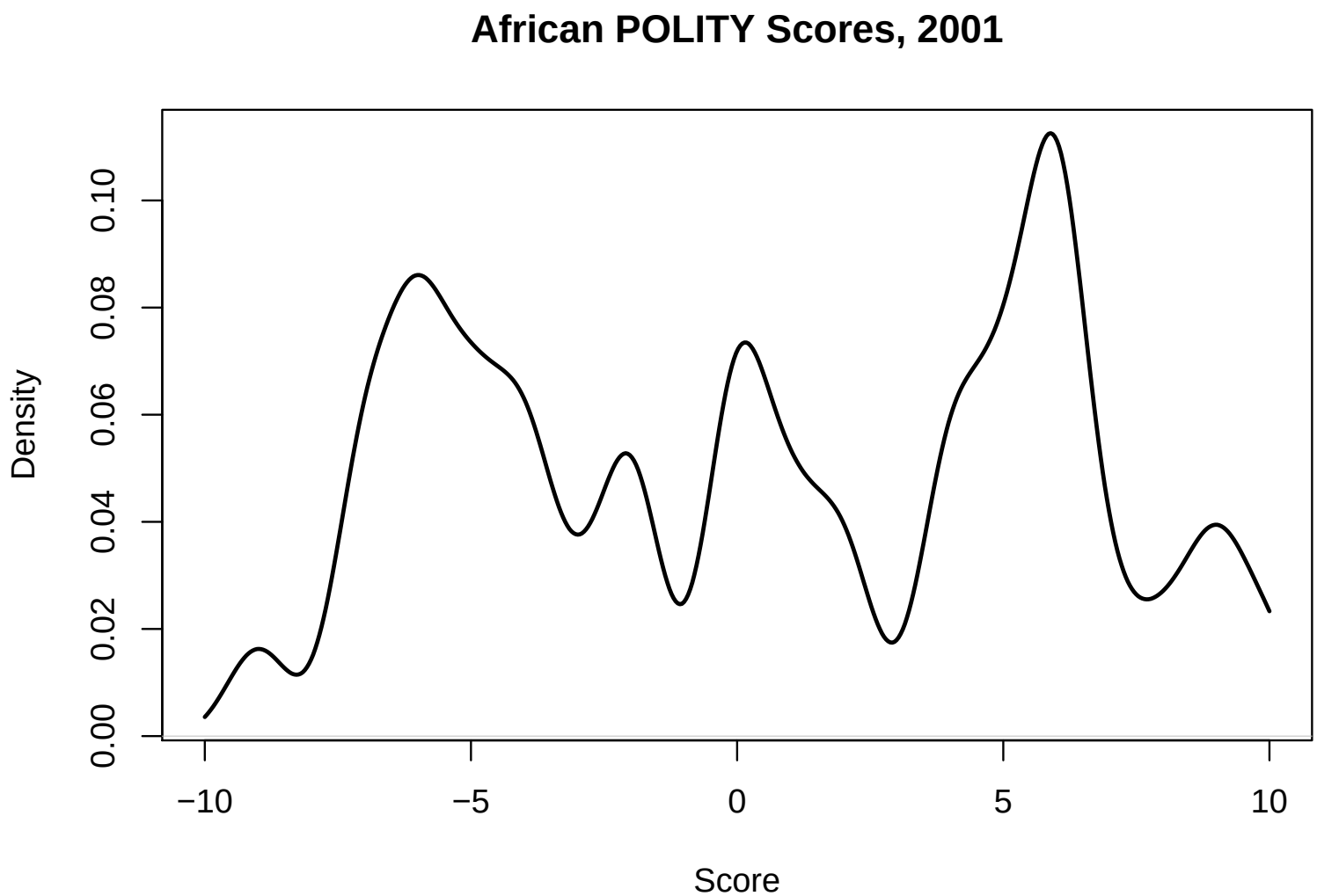
Density Plots

- "Super-Histograms"
- Details are complicated...
- "Smooth" frequencies across different values of the variable
- Can show broad patterns or fine details

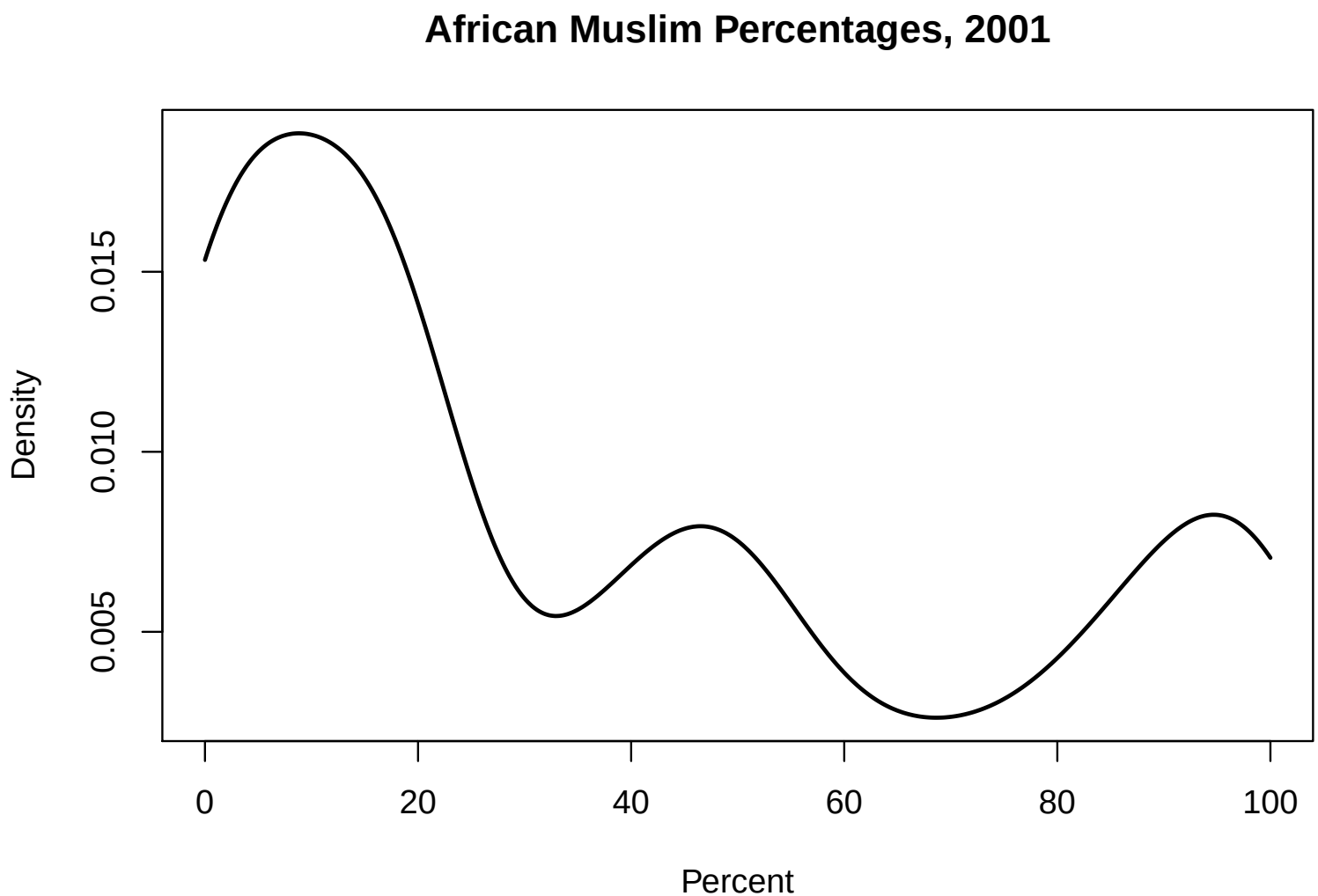
Density Plot: POLITY Scores



POLITY Density: Narrower Bandwidth



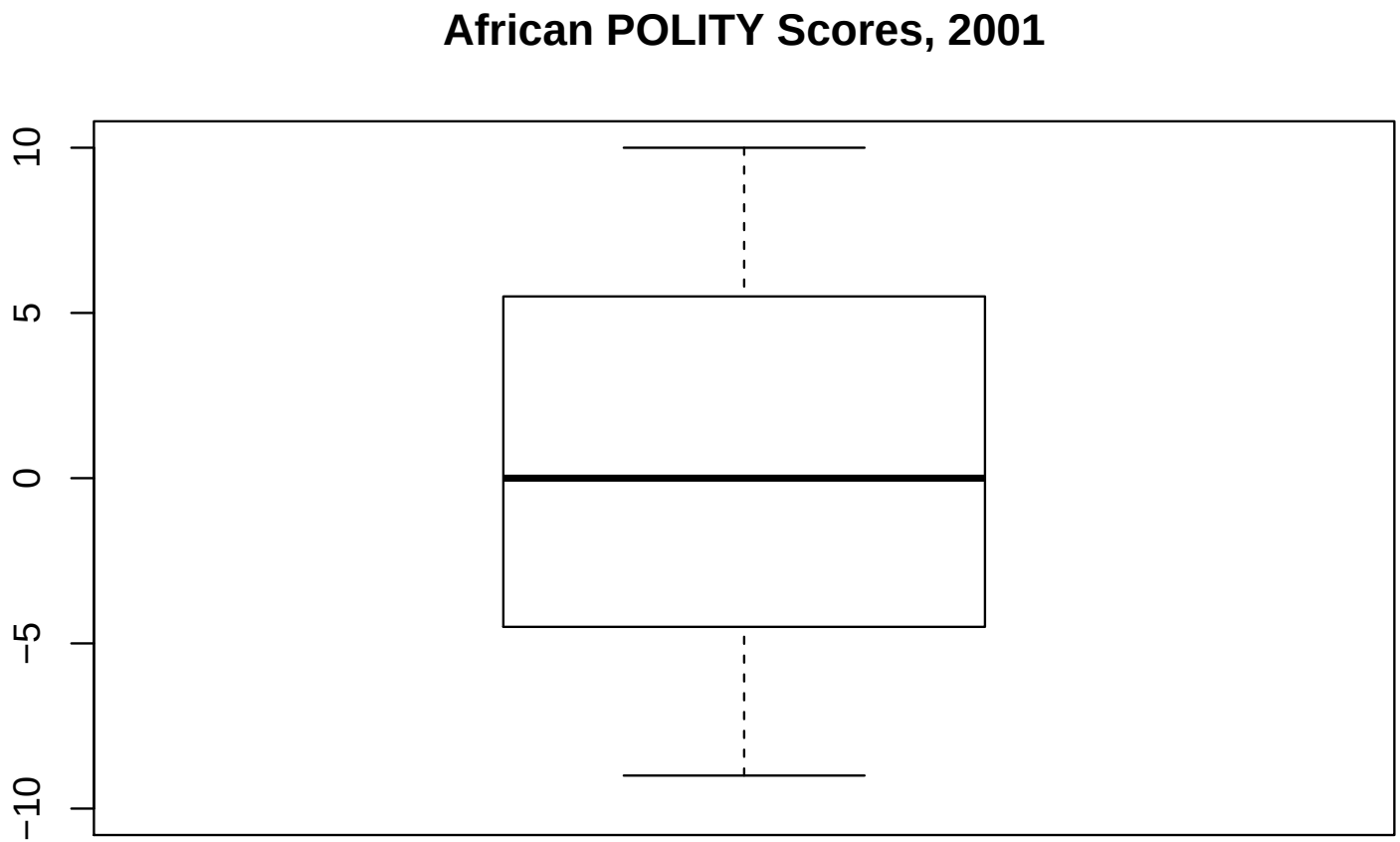
Density Plot: Muslim Percentage



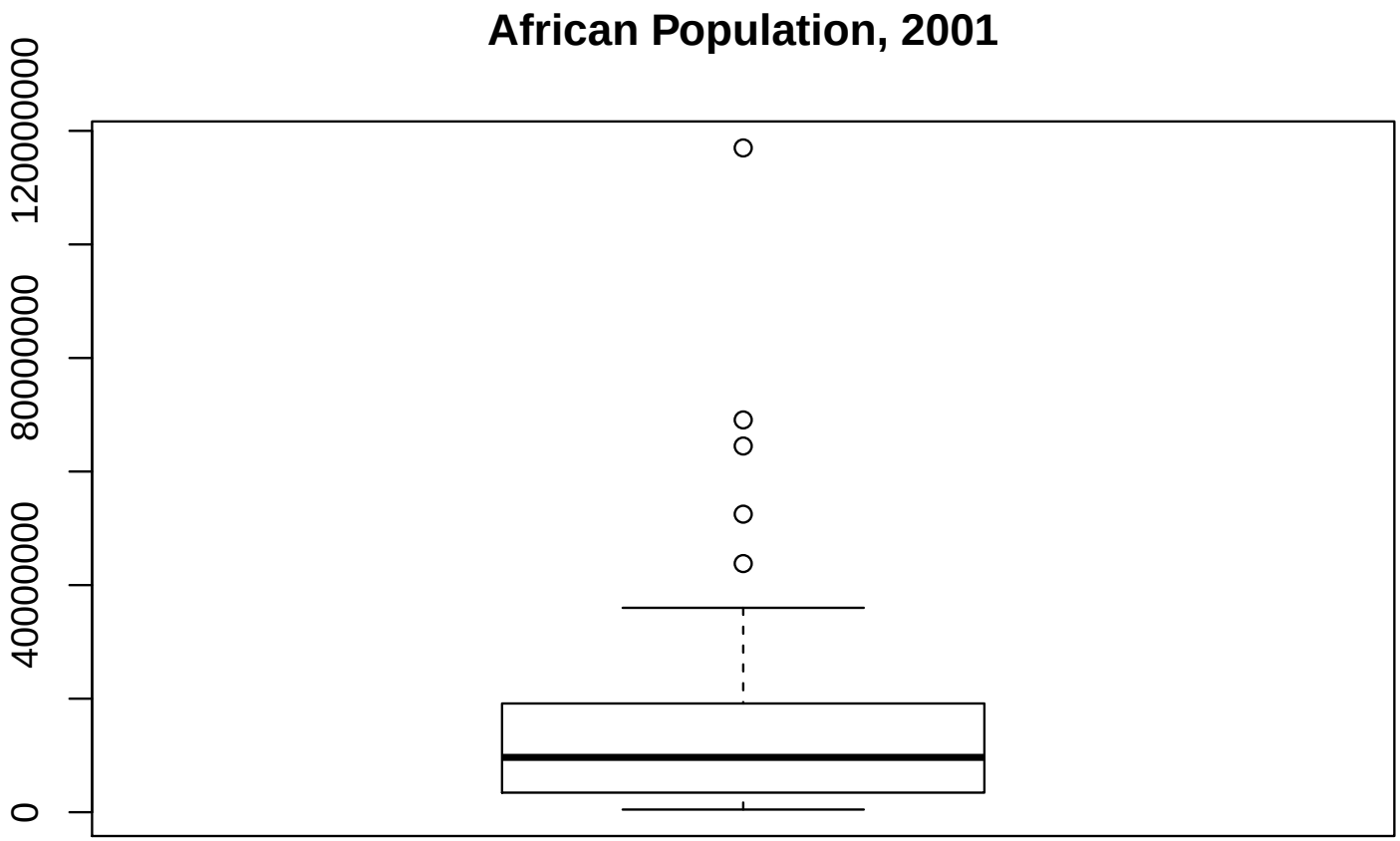
Boxplots

- a/k/a "box-and-whisker" plot
- Imagine looking "down on" a histogram / density plot
- Components:
  - Bar = median
  - Box = *interquartile range* ("IQR;" values from the 25th to the 75th percentile of the data)
  - "Whiskers" = the "fence" (either the minimum & maximum, or 1.5 $\times$ IQR)
  - Points denote "outliers."

Simple Boxplot: POLITY scores (no outliers)



Boxplot with Outliers: Population



Comparing Variables with Boxplots