

# EDA Project 3B -Gapminder UD651 PS-5

*Curtis O'Neal*

*May 5, 2016*

\*\*\*\*\* Gapminder dataset \*\*\*\*\* The Gapminder website contains over 500 data sets with information about the world's population. Your task is to continue the investigation you did at the end of Problem Set 4 or you can start fresh and choose a different data set from Gapminder.

In your investigation, examine 3 or more variables and create 2-5 plots that make use of the techniques from Lesson 5. \*\*\*\*\*

Using Indicator\_BMI male ASM.xlsx

Data set is titled "BMI male, age standardized mean"

The mean BMI (Body Mass Index) of the male population, counted in kilogram per square meter; this mean is calculated as if each country has the same age composition as the world population.

Primary Source: School of Public Health, Imperial College London <http://www.imperial.ac.uk/> MRC-HPA Centre for Environment and Health <http://www.imperial.ac.uk/medicine/globalmetabolics/> Secondary Source: Uploaded to Gapminder 8/2 2011 Downloaded from <http://www.gapminder.org/data/> 5-2-2016

A quick Tour of the Data

```
## [1] "Variable Names"
```

```
## [1] "Country" "X1980" "X1981" "X1982" "X1983" "X1984" "X1985"
## [8] "X1986" "X1987" "X1988" "X1989" "X1990" "X1991" "X1992"
## [15] "X1993" "X1994" "X1995" "X1996" "X1997" "X1998" "X1999"
## [22] "X2000" "X2001" "X2002" "X2003" "X2004" "X2005" "X2006"
## [29] "X2007" "X2008"
```

```
## [1] "Number of Countrys"
```

```
## [1] 199
```

```
## [1] "Rows vs Variables- data set is in wide format"
```

```
## [1] 199 30
```

```
## [1] "199 Unique BMI Values"
```

```
## [1] 20.62058 26.44657 24.59620 27.63048 22.25083 25.76602 27.50170
## [8] 25.35542 27.56373 26.46741 25.65117 27.24594 27.83721 20.39742
## [15] 26.38439 26.16443 26.75915 27.02255 22.41835 28.41894 22.82180
## [22] 24.43335 26.61163 22.12984 25.78623 27.31079 24.18179 26.54286
## [29] 21.27157 21.50291 20.80496 23.68173 27.45210 23.51522 20.99095
## [36] 21.48569 27.01542 22.92176 24.94041 22.06131 19.86692 21.87134
## [43] 32.67440 26.47897 22.56469 26.59629 25.06867 27.41899 27.90524
## [50] 26.13287 23.38403 24.57270 25.19668 25.58841 26.73243 26.36751
## [57] 23.76640 20.88509 26.26446 20.24700 26.53078 26.73339 25.85329
## [64] 30.86752 24.07620 21.65029 25.54942 27.16509 22.84247 26.33786
```

```

## [71] 26.01359 25.17988 25.29947 22.52449 21.64338 23.68465 23.66302
## [78] 25.10872 25.05747 27.11568 27.20687 20.95956 21.85576 25.31003
## [85] 26.71017 27.65325 27.13151 26.48020 24.00421 23.50004 27.47362
## [92] 26.29078 21.59258 29.23840 22.01726 23.98950 29.17211 24.74743
## [99] 21.07931 26.45693 27.20117 21.90157 21.89537 26.54164 26.86102
## [106] 27.43404 25.71382 26.34473 21.40347 22.03468 24.73069 23.21991
## [113] 21.78881 27.68361 29.37337 22.62295 25.15669 27.42468 28.10315
## [120] 24.23690 24.88385 26.55412 25.63182 21.93536 21.44932 22.65008
## [127] 33.89634 20.76344 26.01541 27.70558 27.76893 25.77291 21.21958
## [134] 23.03322 26.93424 26.24109 22.29914 30.37757 26.26959 25.01506
## [141] 25.54223 24.77041 22.87263 26.67380 26.68445 28.37804 28.13138
## [148] 25.41069 26.01131 22.55453 28.22986 24.65176 25.44121 30.42475
## [155] 23.51233 27.88432 21.92743 26.51495 25.56236 22.53139 23.83996
## [162] 26.92717 27.43983 27.15988 21.96917 26.85538 27.49975 21.96671
## [169] 22.40484 25.49887 23.16969 26.37629 26.20195 26.91969 24.44939
## [176] 23.77966 22.47792 23.00803 20.59082 21.87875 30.99563 26.39669
## [183] 25.15699 26.70371 25.24796 22.35833 25.42379 28.05359 27.39249
## [190] 28.45698 26.39123 25.32054 26.78926 27.44500 20.91630 26.57750
## [197] 24.44157 20.68321 22.02660

## [1] "Adjust the Data Set to Tall format"

## Using Country as id variables

## [1] "Now has these Rows, and 3 Variables"

## [1] 5771      3

## [1] "3 Variables Named-"

## [1] "Country"      "Year"          "Ave_Male_BMI"

## [1] "Summary Statistics of Ave Male BMI"

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      19.01  21.87   24.42   24.01   25.69   33.90

## [1] "For Summary Graphing, Values of BMI Rounded Down Could Be Taken"

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      19.0   21.0   24.0   23.5   25.0   33.0

## [1] "Giving us fewer values of BMI, that would fit on a graph"

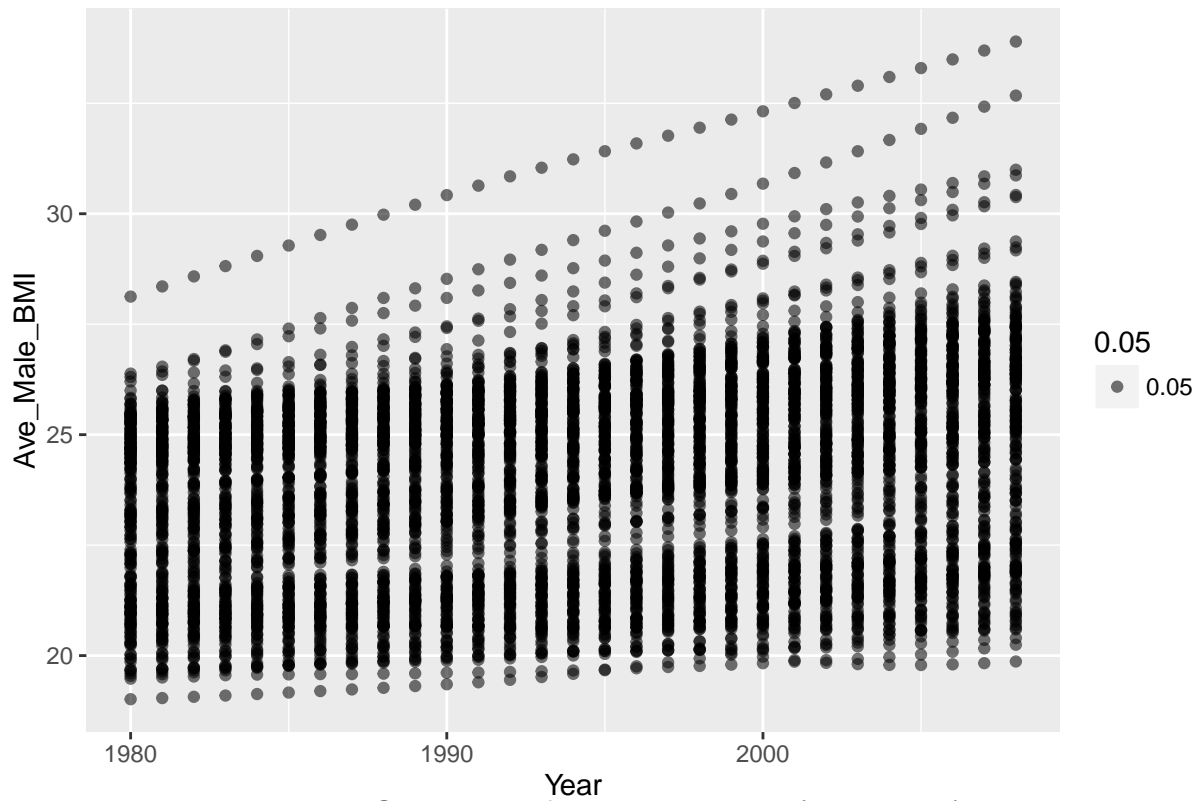
## [1] 15

## [1] "Clean up the data, check the structure and values-"

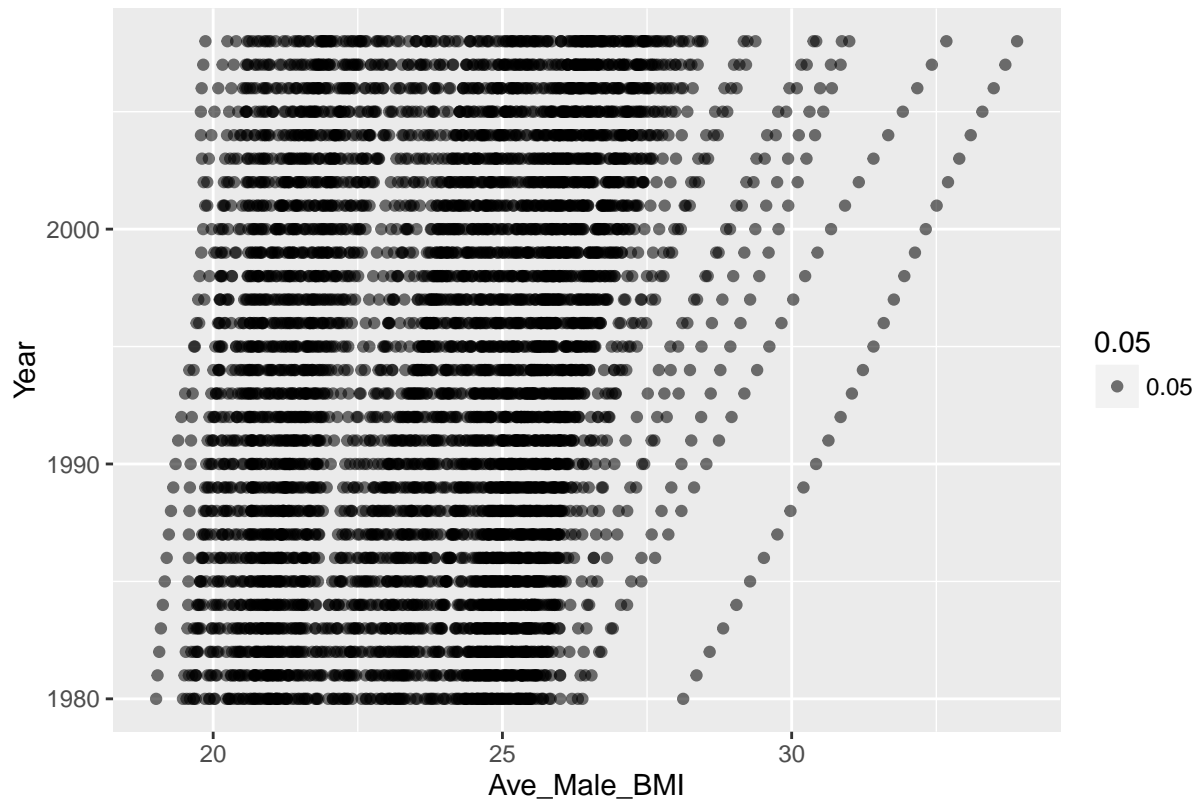
## 'data.frame':   5771 obs. of  4 variables:
##  $ Country      : Factor w/ 199 levels "Afghanistan",...: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Year         : num  1980 1980 1980 1980 1980 1980 1980 1980 1980 1980 ...
##  $ Ave_Male_BMI: num  21.5 25.2 22.3 25.7 20.9 ...
##  $ Floor_BMI   : num  21 25 22 25 20 23 25 23 24 24 ...

```

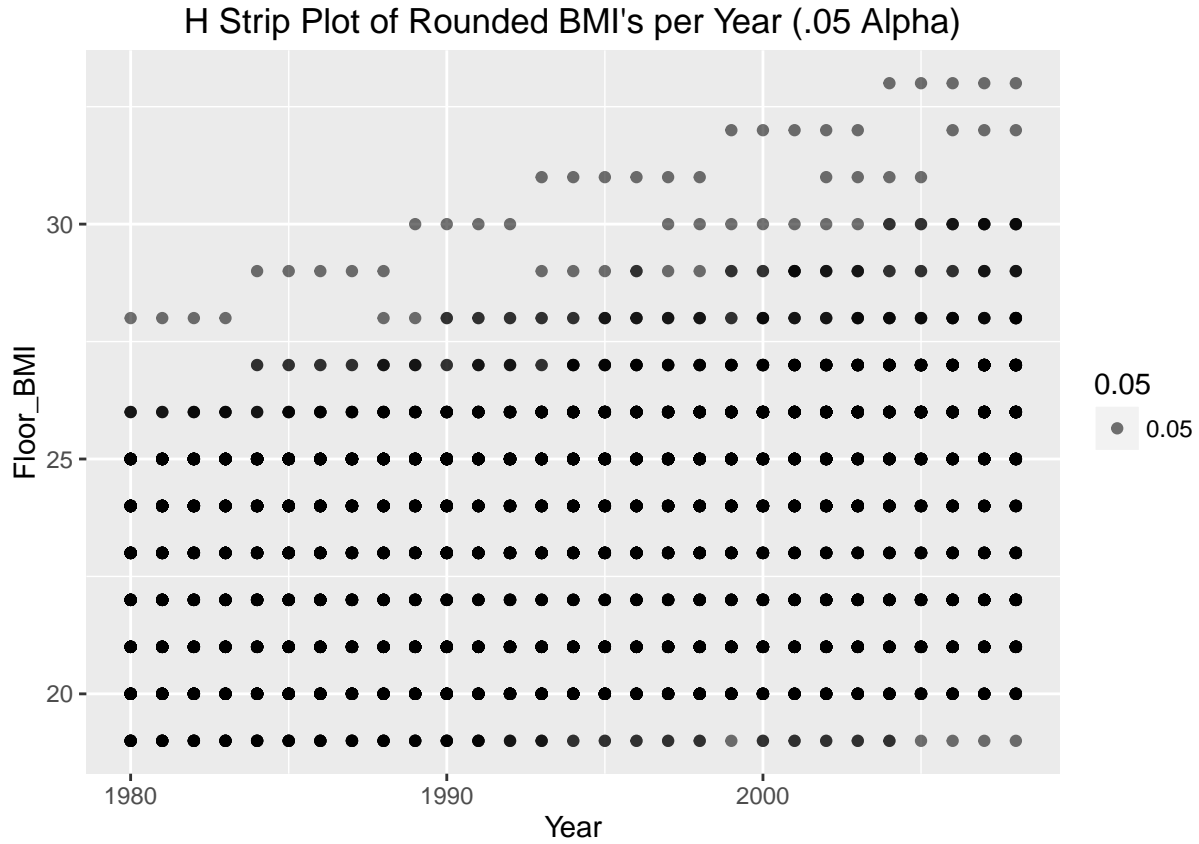
Vertical Strip Plot of BMI's per Year (.05 Alpha)



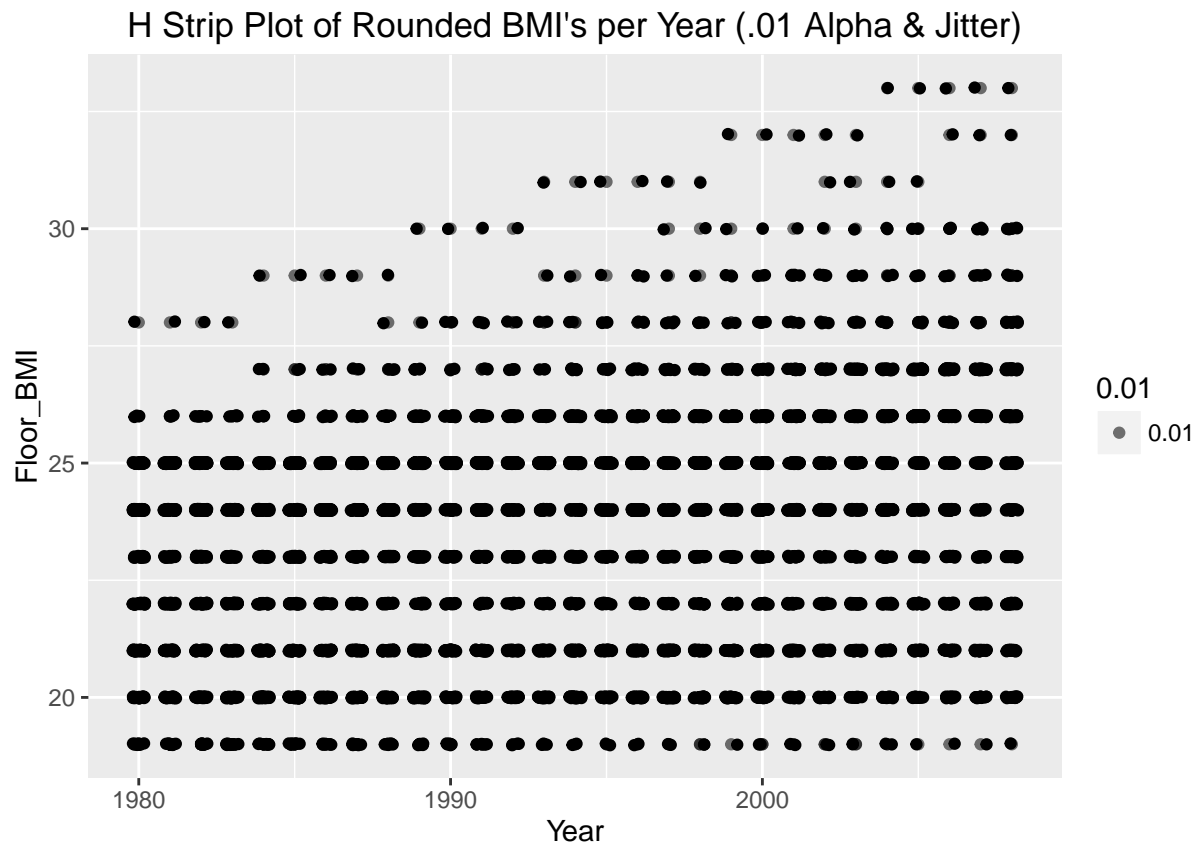
Horizontal Strip Plot of BMI's per Year (.05 Alpha)



```
## [1] "These two graphs start to show us the general change, \nbut not the effect of time for specific"
```

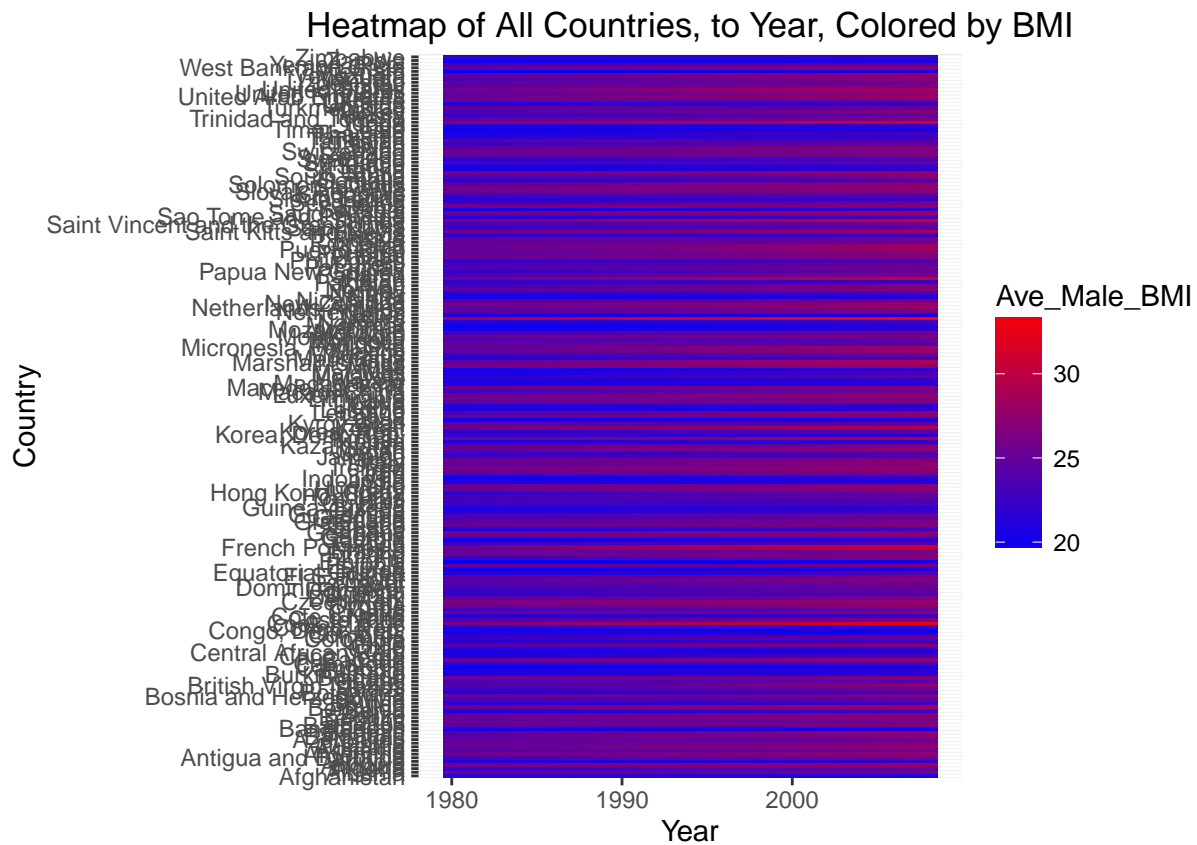


```
## [1] "Rounding removes the clutter but doesn't give a sense of \ntrends, rather than a general visual"
```



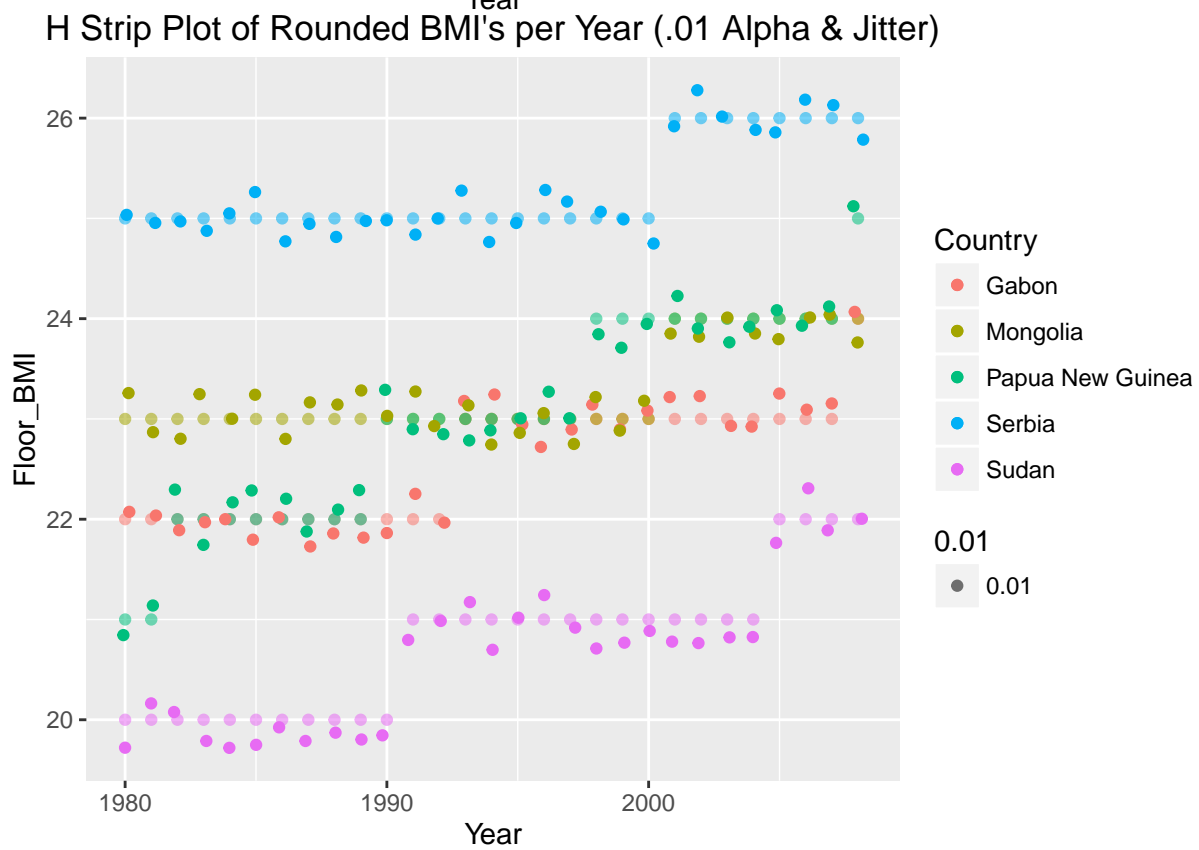
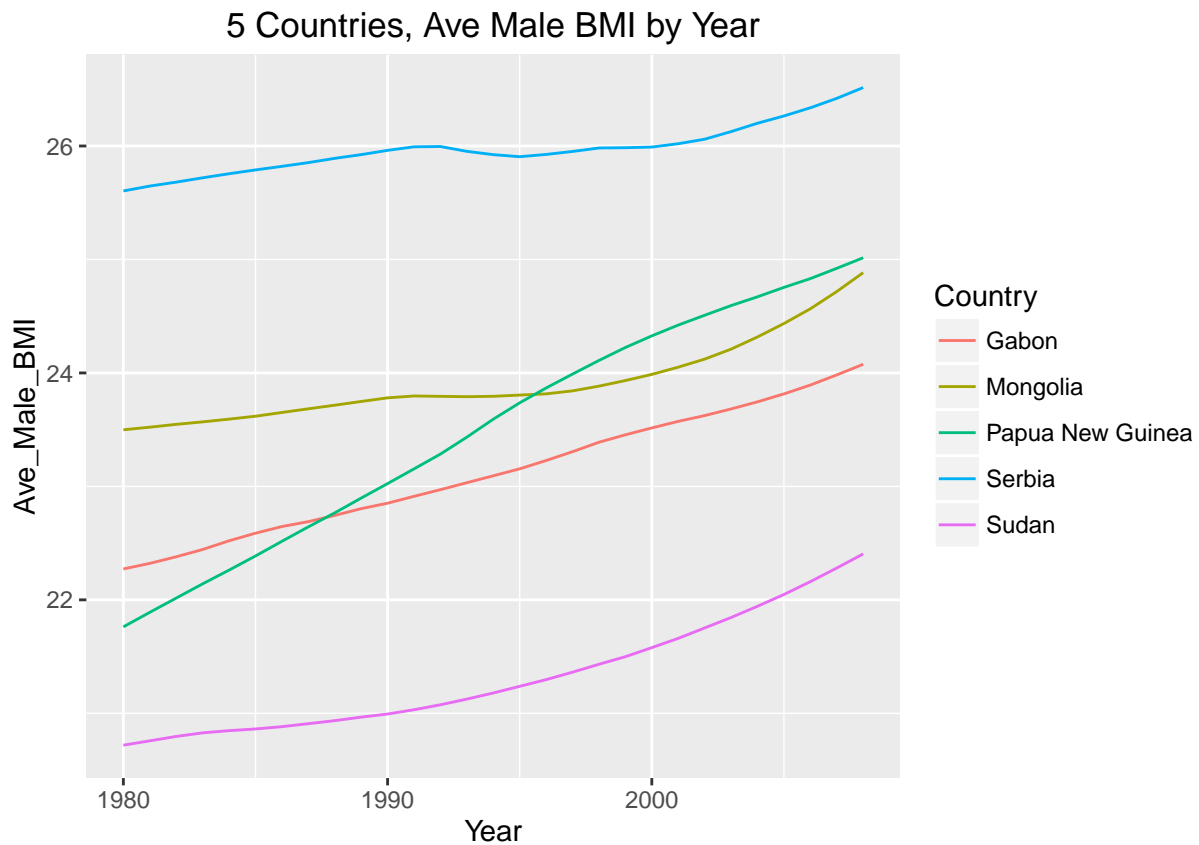
```
## [1] "Transparency and jittering doesn't provide as much of a \nsense of the rarity of particular va
```

```
## [1] "For this much data we might try the heatmap or tile used in genomic data-"
```

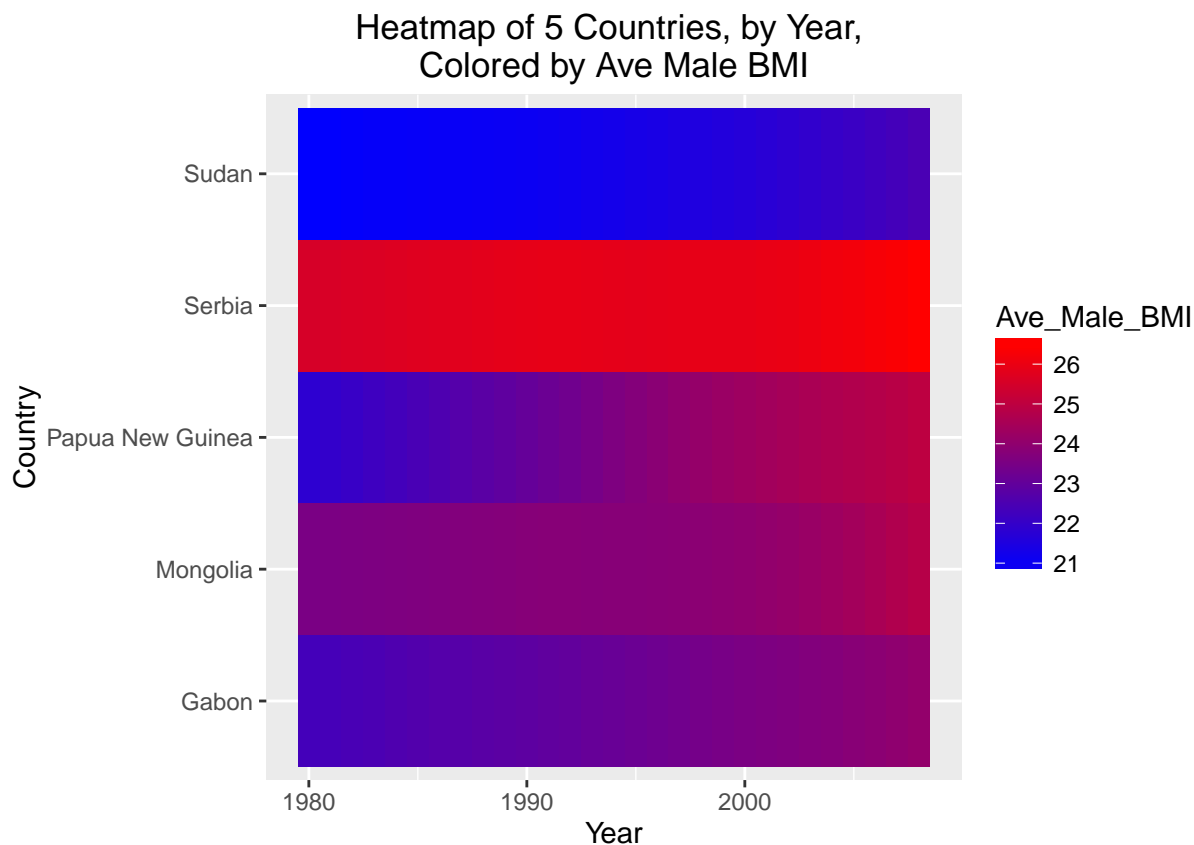


```
## [1] "For this we'd need a better way to identify the countries in \na systemitized way, but we do se
```

```
## [1] "To make sense of some specific trends to a limited subset  with a random sample of \n5 countries
```



## [1] "This plot is specifically interesting due to the large jumps from \n year to year in some countr



```
## [1] "The joupn in Serbia appears more gentil in the heatmap and \nthe line graph versions."
```

```
## [1] "Summary Statistics Using All Countries Values-"
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  19.01   21.87   24.42   24.01   25.69   33.90
```

```
## [1] "Summary Statistics Using Only te 5 Subset Countries-"
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  20.72   22.39   23.62   23.57   24.59   26.51
```

```
## [1] "A Table by Subset Country of Summary Statistics-"
```

```
##
##      Gabon      Mongolia Papua New Guinea      Serbia
##      29         29         29         29
##      Sudan
##      29
```

```
## m_dat2$Country: Gabon
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  22.27   22.69   23.09   23.13   23.57   24.08
```

```
## -----
```

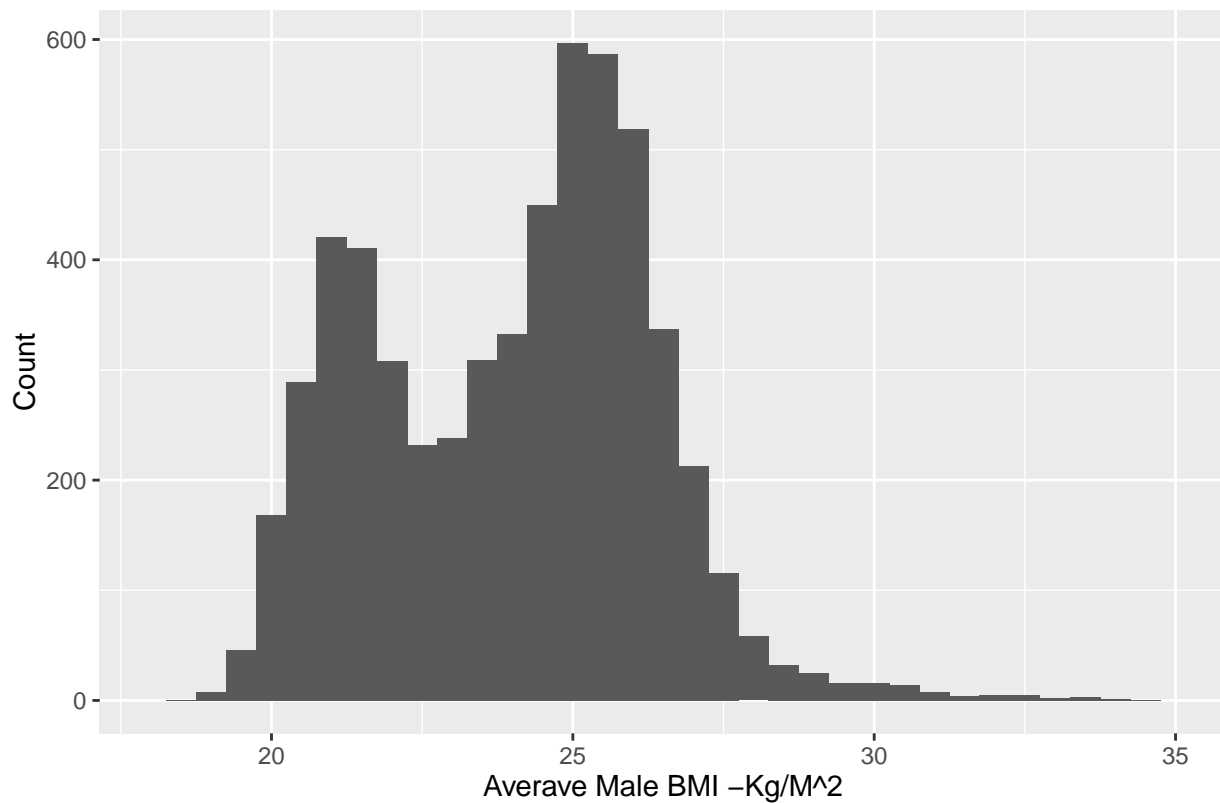
```
## m_dat2$Country: Mongolia
```



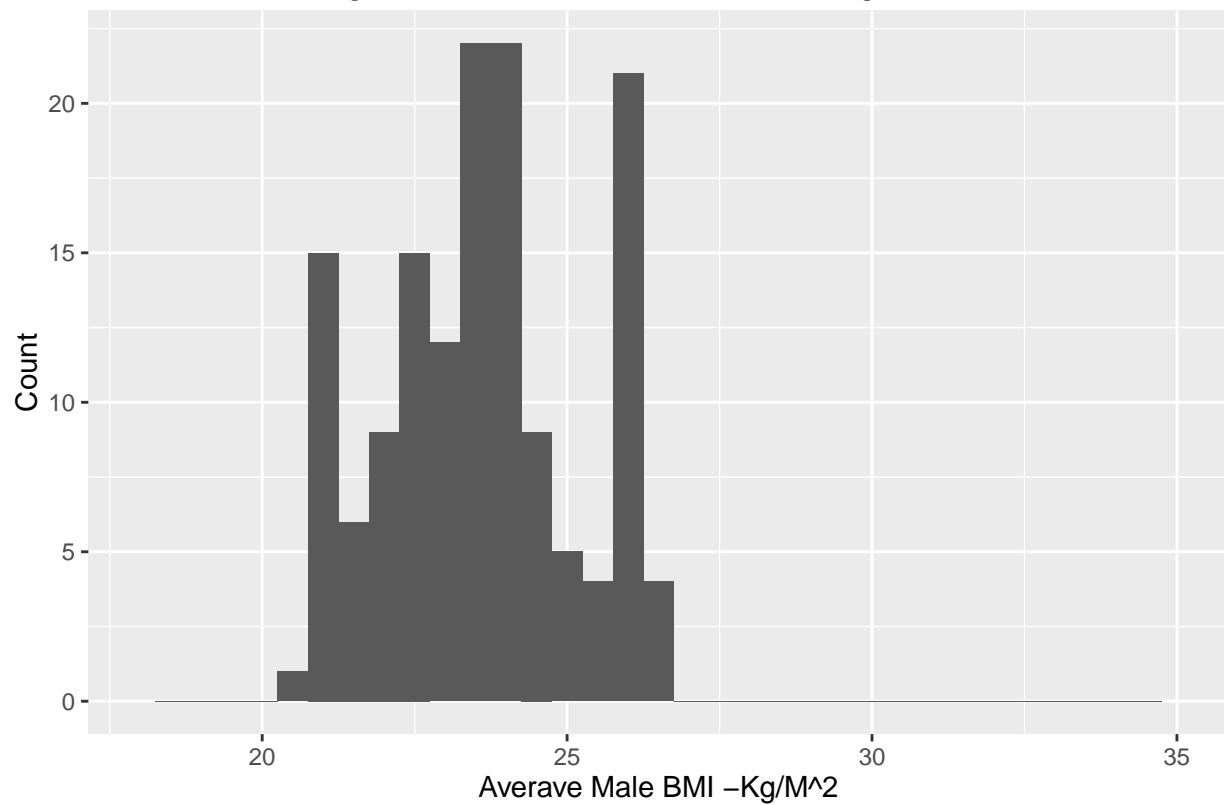
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      23.50  23.68   23.80   23.92  24.05   24.88
## -----
## m_dat2$Country: Papua New Guinea
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      21.76  22.64   23.59   23.51  24.42   25.02
## -----
## m_dat2$Country: Serbia
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      25.60  25.85   25.95   25.97  26.02   26.51
## -----
## m_dat2$Country: Sudan
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      20.72  20.91   21.18   21.32  21.66   22.40
```

Plotting the Data Individually - redoing some of the previous plots with the subset countries.

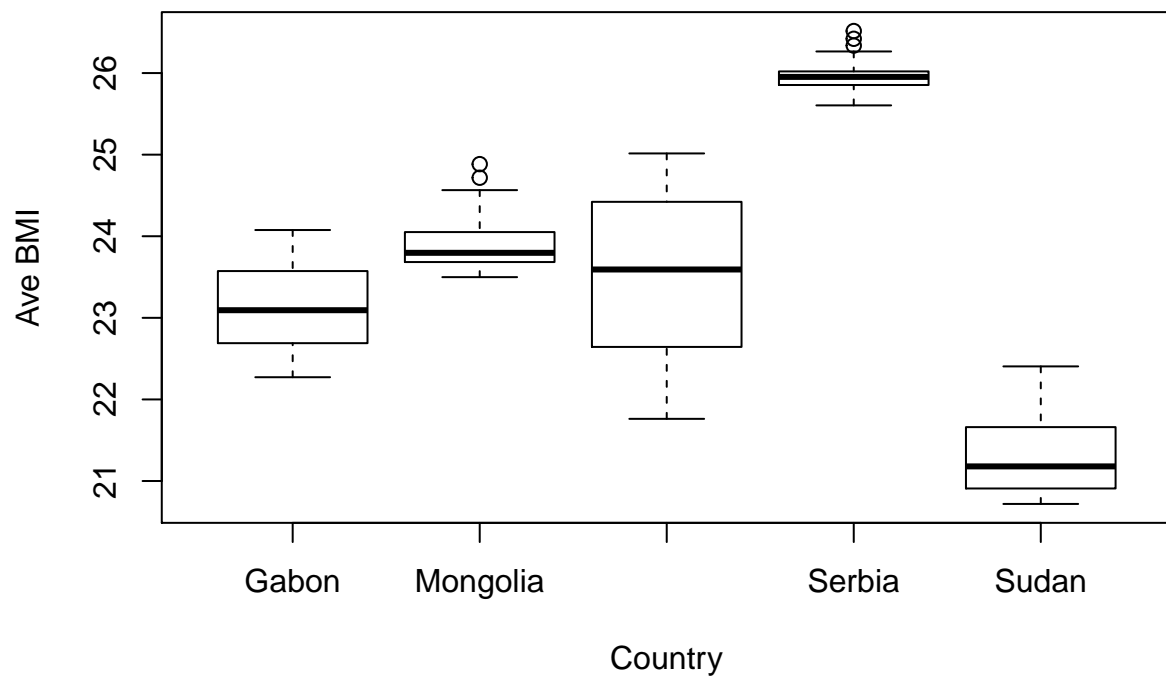
### Histogram of all BMI Values in OriginalDataset



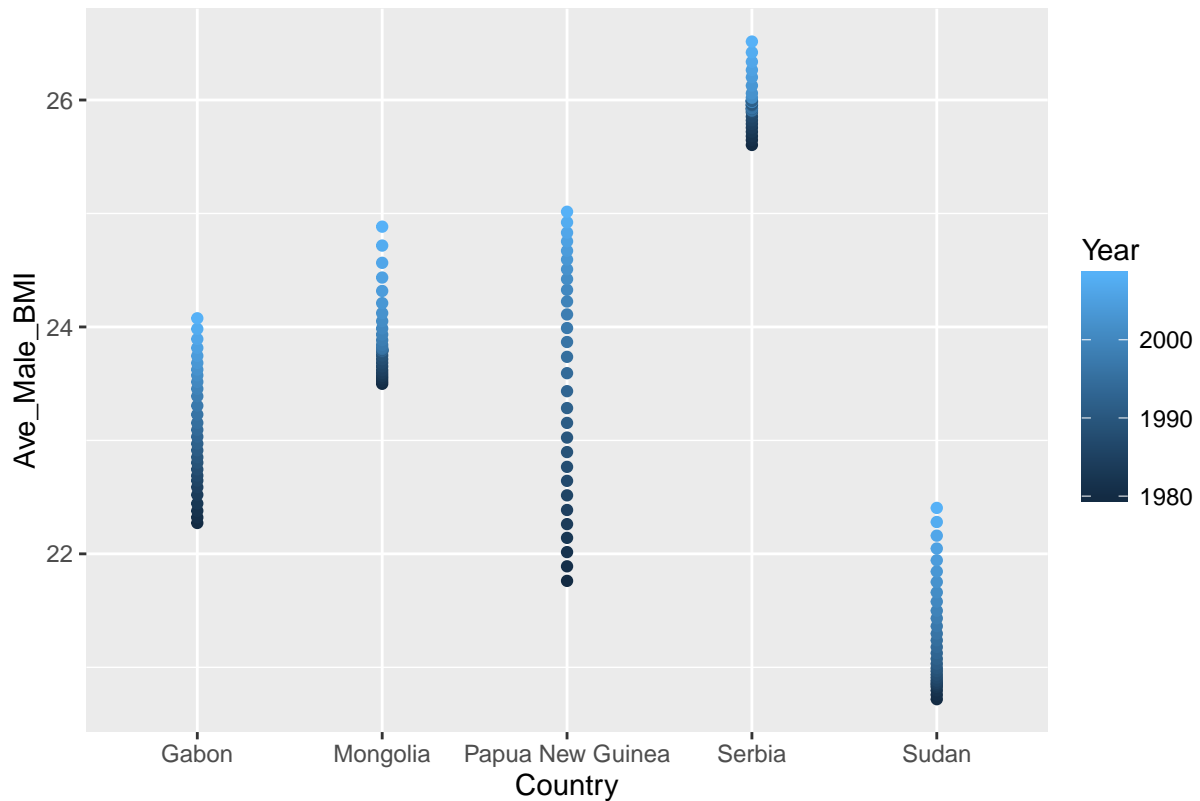
Histogram of Subset BMI Values in OriginalDataset



**Ave Male BMI, by Selected Counties,1980–2008**



Vertical Strip Plot of the Average Male BMI Data, by Year



Horizontal Strip Plot of the Average Male BMI Data, by Year

