StatR 101: Fall 2012

Homework 2

Rod Doe

Saturday, October 6, 2012

# Analysis of table tennis appeal:

## Prediction of appeal of table tennis by gender:

I predict that that males will be more interested in ping pong than will females in data undifferentiated by nationality.

## Table to summarize interest in table tennis:

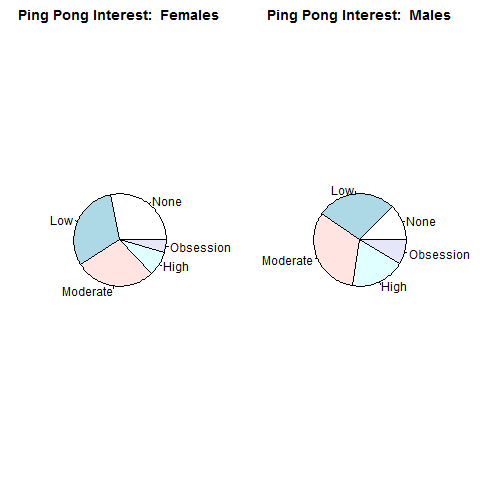
None Low Moderate High Obsession

PingPong.All 19 27 28 13 6

PingPong.Females 13 14 13 4 2

PingPong.Males 6 13 15 9 4

## Side-by-side pie charts of interest in Ping Pong:



## 2 x 5 matrix summarizing the proportional distribution of male/female students in each ping pong interest:

> PingPongMatrix

None Low Moderate High Obsession

PingPong.All 19 27 28 13 6

PingPong.Females 13 14 13 4 2

PingPong.Males 6 13 15 9 4

Females = PingPongMatrix[2,] / sum(PingPongMatrix[2,])

Males = PingPongMatrix[3,] / sum(PingPongMatrix[3,])

M1 = rbind(Females, Males)

M1

None Low Moderate High Obsession

Females 0.2826087 0.3043478 0.2826087 0.08695652 0.04347826

Males 0.1276596 0.2765957 0.3191489 0.19148936 0.08510638

## 5 x 2 matrix summarizing the proportion for each response of male/female respondents such that Pmale,i + Pfemale,I = 1.

> PingPongMatrix

None Low Moderate High Obsession

PingPong.All 19 27 28 13 6

PingPong.Females 13 14 13 4 2

PingPong.Males 6 13 15 9 4

> M2 = cbind(PingPongMatrix[2,]/PingPongMatrix[1,], PingPongMatrix[3,]/PingPongMatrix[1,])

> M2

[,1] [,2]

None 0.6842105 0.3157895

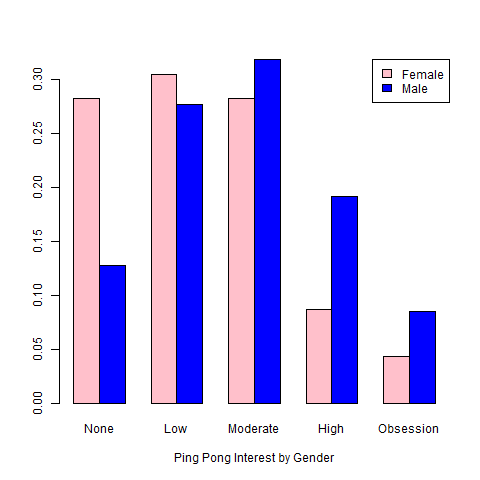
Low 0.5185185 0.4814815

Moderate 0.4642857 0.5357143

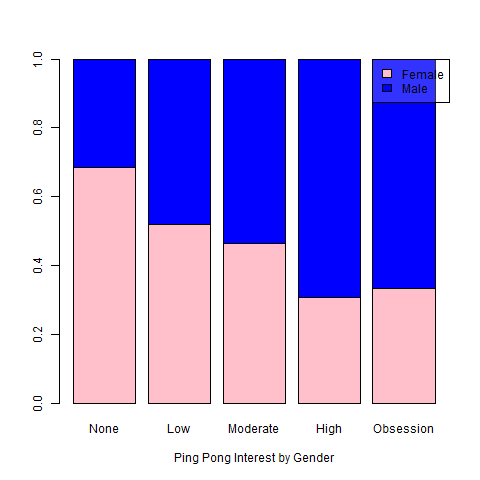
High 0.3076923 0.6923077

Obsession 0.3333333 0.6666667

## Bar plot of ping pong interest by gender



The quality of this graphic competes favorably with Excel. And I am a major Excel fan boy.



## Discussion

My prediction is supported by data. The kinetic college males show a higher level of interest in ping pong. The females evidently have better things to do.

The comparison of depictions is interesting. Numbers are interesting, but graphics turn data into knowledge. IMHO, the last depiction is the most informative. One thing that troubles me is the fact that we have nearly identical numbers of males and females. Had this been a military academy where there are far more males than females, then a treemap (shows population size and intensity) would be a better choice.

# Analysis of World Data

## Top 10 and Lowest 10 Per-Capita GDP Values by Country

Country PerCapitaGDP

1 Congo, Democratic Republic of 329

2 Liberia 396

3 Burundi 412

4 Zimbabwe 436

5 Eritrea 683

6 Central African Republic 747

7 Niger 761

8 Sierra Leone 810

9 Malawi 821

10 Togo 863

11 Netherlands 40973

12 Switzerland 41950

13 Hong Kong 45944

14 United States 46860

15 United Arab Emirates 47439

16 Brunei 48333

17 Norway 51959

18 Singapore 56694

19 Luxembourg 81466

20 Qatar 88222

## Top 10 and Lowest 10 Population Density by Country

Country Density

1 Macau 1.856000e+04

2 Monaco 1.750000e+04

3 Dominica 1.248844e+04

4 Singapore 7.150282e+03

5 Hong Kong 6.428986e+03

6 Gibraltar 4.906833e+03

7 Bahrain 1.628755e+03

8 Malta 1.321544e+03

9 Bermuda 1.195667e+03

10 Sint Maarten 1.100853e+03

11 Suriname 3.204737e+00

12 Botswana 3.092952e+00

13 Iceland 3.091767e+00

14 Australia 2.954077e+00

15 Namibia 2.533968e+00

16 Western Sahara 1.996241e+00

17 Mongolia 1.804808e+00

18 Pitcairn Islands 1.063830e+00

19 Falkland Islands 2.464471e-01

20 Greenland 2.606175e-02

Note: Vatican City reports an area of 0, yielding an infinite population density. His Infallible Holiness has been notified.

Now it is necessary to add columns Population, Area, and PercentWater to the above table (implemented in data frame dfPopDens). I found a sweet way to do this. To get the data for the Area column, this works:

ar = CountryData$Area[CountryData$Country %in% dfPopDens$Country]

Repeat for Country (co), Population (po), and PercentWaterCoverage (pw), and then create the data frame with this:

dfc = data.frame(list(Country=co, Population=po, Area=ar, PercentWaterCoverage=pw))

Country Population Area PercentWaterCoverage

1 Macau 531000 266000 0.00

2 Monaco 3000 12173 0.00

3 Dominica 50 47 0.00

4 Singapore 37429 34 0.00

5 Hong Kong 2088669 824268 0.12

6 Gibraltar 1800098 582000 2.58

7 Bahrain 9378818 751 0.72

8 Malta 2822900 1564100 0.68

9 Bermuda 1234596 758 0.00

10 Sint Maarten 525000 163820 4.77

11 Suriname 56452 2166086 NA

12 Botswana 318452 103000 2.67

13 Iceland 22722835 7692024 0.76

14 Australia 64566 54 0.00

15 Namibia 29441 6 0.00

16 Western Sahara 417608 316 0.00

17 Mongolia 5076700 710 1.43

18 Pitcairn Islands 35000 2 0.00

19 Falkland Islands 556800 30 0.00

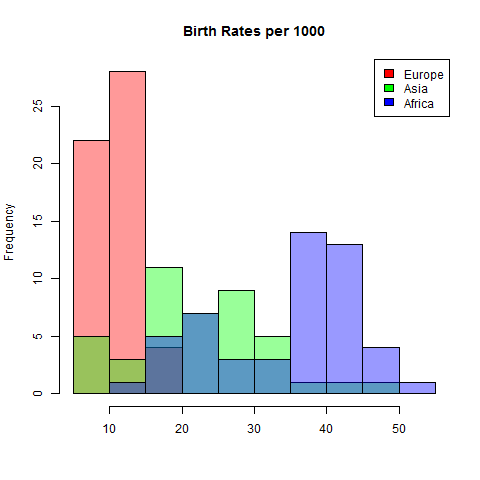
20 Greenland 7097600 1104 4.53

## Frequency Histogram of Birth Rates in Europe, Asia, and Africa

eu = CountryData$Birthrate[CountryData$Continent == "Europe"]

as = CountryData$Birthrate[CountryData$Continent == "Asia"]

af = CountryData$Birthrate[CountryData$Continent == "Africa"]



hist(eu, breaks=c(5,10,15,20,25,30,35,40,45,50,55), col=rgb(1,0,0,0.4), main="Birth Rates per 1000", xlab="")

hist(as, add=TRUE, col=rgb(0,1,0,0.4))

hist(af, add=TRUE, col=rgb(0,0,1,0.4))

legend("topright", fill=c(rgb(1,0,0), rgb(0,1,0), rgb(0,0,1)), legend=c("Europe", "Asia", "Africa"))

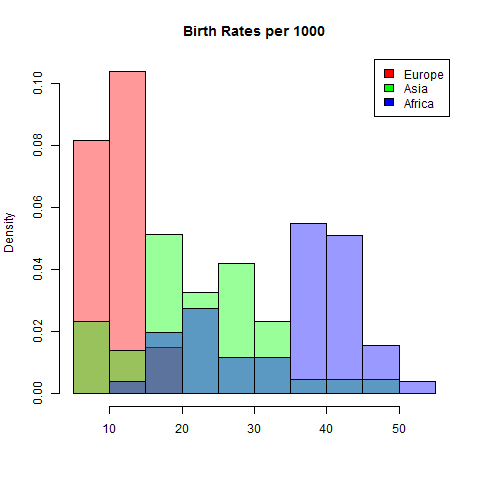
## Density Histogram of Birth Rates in Europe, Asia, and Africa

hist(eu, freq=FALSE, breaks=c(5,10,15,20,25,30,35,40,45,50,55), col=rgb(1,0,0,0.4), main="Birth Rates per 1000", xlab="")

hist(as, freq=FALSE, add=TRUE, col=rgb(0,1,0,0.4))

hist(af, freq=FALSE, add=TRUE, col=rgb(0,0,1,0.4))

legend("topright", fill=c(rgb(1,0,0), rgb(0,1,0), rgb(0,0,1)), legend=c("Europe", "Asia", "Africa"))



## Summary of Birth Rate Data

The data are not normally distributed. The data are bimodal, with the low frequency Europeans constituting the lower birth rates, and with Africa contributing to the high birth rate.