

Lab1.Assignment_Emily-Chu

Code ▾

This is an R Markdown (<http://rmarkdown.rstudio.com>) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

Hide

```
plot(cars)
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

Data frame for this assignment: country

Now, lets proceed to the lab which will be conducted in this notebook. Please show all of your code and output and upload the HTML file notebook as your assignment. You must write up your results independently.

1. List the level of measurement for each of the following variables:
 - a. hdicat: Nominal
 - b. urban2004: Interval-ratio
 - c. internet03: Interval-ratio
 - d. educ04: Interval-ratio
2. What is the average public expenditure on education (% of GDP) in 2004?

4.996

Hide

```
summary(country)
```

case	country	hdirank	hdicat
Min. : 1	Length:193	Min. : 1	Length:193
1st Qu.: 49	Class :character	1st Qu.: 45	Class :character
Median : 97	Mode :character	Median : 89	Mode :character
Mean : 97		Mean : 89	
3rd Qu.:145		3rd Qu.:133	
Max. :193		Max. :177	
		NA's :16	

pop2004	urban2004	physicians	mort2004
Min. : 0.10	Min. : 9.70	Min. : 2.00	Min. : 2.00
1st Qu.: 2.15	1st Qu.: 34.80	1st Qu.: 25.25	1st Qu.: 10.00
Median : 7.25	Median : 56.60	Median : 107.50	Median : 24.50
Mean : 34.30	Mean : 54.52	Mean : 170.91	Mean : 41.74
3rd Qu.: 22.25	3rd Qu.: 72.40	3rd Qu.: 250.00	3rd Qu.: 68.00
Max. :1308.00	Max. :100.00	Max. :4735.00	Max. :165.00
NA's :7		NA's :3	NA's :1

tele2004	cell2003	internet03	gdp
Min. : 1.00	Min. : 2.0	Min. : 1.0	Min. : 90.0
1st Qu.: 40.25	1st Qu.: 77.0	1st Qu.: 24.0	1st Qu.: 632.8
Median :152.50	Median : 289.0	Median : 73.0	Median : 2342.5
Mean :212.94	Mean : 386.8	Mean :171.9	Mean : 7992.6
3rd Qu.:329.75	3rd Qu.: 652.0	3rd Qu.:265.0	3rd Qu.: 7875.8
Max. :710.00	Max. :1184.0	Max. :788.0	Max. :70295.0
NA's :51	NA's :34	NA's :12	NA's :13

inequality	educ04	milit2004	debt04
Min. : 3.300	Min. : 0.600	Min. : 0.000	Min. : 0.10
1st Qu.: 8.125	1st Qu.: 3.550	1st Qu.: 1.100	1st Qu.: 1.80
Median : 12.500	Median : 4.700	Median : 1.700	Median : 4.00
Mean : 21.060	Mean : 4.996	Mean : 2.186	Mean : 5.56
3rd Qu.: 19.275	3rd Qu.: 5.950	3rd Qu.: 2.600	3rd Qu.: 7.80
Max. :168.100	Max. :16.000	Max. :12.000	Max. :30.40
NA's :67	NA's :66	NA's :60	NA's :60

cde2003	lifef	lifem
Min. : 0.100	Min. :31.3	Min. :31.30
1st Qu.: 0.200	1st Qu.:57.3	1st Qu.:54.85
Median : 0.300	Median :72.8	Median :67.10
Mean : 1.166	Mean :67.7	Mean :63.29
3rd Qu.: 1.000	3rd Qu.:77.7	3rd Qu.:71.70
Max. :23.000	Max. :85.6	Max. :79.00
NA's :108	NA's :14	NA's :14

3. What is the minimum urban population rate (% of total) in 2004?

9.7

Hide

```
min(country$urban2004)
```

```
[1] 9.7
```

4. What is the median for internet users per 1,000 people in 2003?

73.0

Hide

```
summary(country$internet03)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.0	24.0	73.0	171.9	265.0	788.0	12

5. What is the 95th percentile for internet users per 1,000 people in 2003?

629

Hide

```
quantile(country$internet03, 0.95, na.rm=TRUE)
```

95%
629

6. Compare the median and 95th percentile for internet users per 1,000 people in 2003. What does this comparison tell you about the distribution of this variable?

Hide

```
summary(country$internet03)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.0	24.0	73.0	171.9	265.0	788.0	12

Hide

```
quantile(country$internet03, 0.95, na.rm=TRUE)
```

95%
629

It is highly positively skewed with the median much lower than the mean and a few very high number of internet users per 1,000 people with the 95th percentile tailing to 629 internet users per 1,000 people.

7. Describe hdicat. Use descriptive statistics, if appropriate.

hdicat is a nominal level variable that measures if a country is classified as low, medium or high development. Around 32 (18%) of countries in the dataset are classified as low development, 88 (49.7%) as medium development and 57 (32%) as high development. The mode for hdicat is "Medium Development".

Hide

```
summary(country$hdicat)
```

Length	Class	Mode
193	character	character

Hide

```
table(country$hdicat)
```

High Development	Low Development	Medium Development
57	32	88

Hide

```
prop.table(table(country$hdicat))
```

High Development	Low Development	Medium Development
0.3220339	0.1807910	0.4971751

Hide

```
names(sort(-table(country$hdicat)))[1]
```

```
[1] "Medium Development"
```

8. Why is it important to use Frequency Tables for nominal and ordinal variables and other descriptive statistics for interval- ratio variables?

It is important to use Frequency Tables for nominal and ordinal variables because median and mean are not typically used at this level of measurement. Calculating mode and frequency is more appropriate. For interval-ratio variables, median, mean and other descriptive statistics are appropriate for the level of measurement.

Submit to Canvas in both HTML file and PDF file