

ASSIGNMENT 1

Siddhartha CBS

2023-09-09

#2 Import the dataset into R

```
mydata=read.csv("C:/Users/Siddhartha/Downloads/world_population.csv")
```

#3 Print out descriptive statistics for a selection of quantitative and categorical variables.

```
summary(mydata)
```

```
##           Rank           CCA3           Country           Capital
##  Min.      : 1.00      Length:234      Length:234      Length:234
## 1st Qu.: 59.25      Class :character      Class :character      Class :character
## Median :117.50      Mode  :character      Mode  :character      Mode  :character
## Mean      :117.50
## 3rd Qu.:175.75
## Max.      :234.00
##
## Continent      X2022.Population      X2020.Population      X2015.Population
## Length:234      Min.      :5.100e+02      Min.      :5.200e+02      Min.      :5.640e+02
## Class :character      1st Qu.:4.197e+05      1st Qu.:4.153e+05      1st Qu.:4.047e+05
## Mode  :character      Median :5.560e+06      Median :5.493e+06      Median :5.307e+06
##                               Mean      :3.407e+07      Mean      :3.350e+07      Mean      :3.173e+07
##                               3rd Qu.:2.248e+07      3rd Qu.:2.145e+07      3rd Qu.:1.973e+07
##                               Max.      :1.426e+09      Max.      :1.425e+09      Max.      :1.394e+09
##
## X2010.Population      X2000.Population      X1990.Population
## Min.      :5.960e+02      Min.      :6.510e+02      Min.      :7.000e+02
## 1st Qu.:3.931e+05      1st Qu.:3.272e+05      1st Qu.:2.641e+05
## Median :4.943e+06      Median :4.293e+06      Median :3.825e+06
## Mean      :2.985e+07      Mean      :2.627e+07      Mean      :2.271e+07
## 3rd Qu.:1.916e+07      3rd Qu.:1.576e+07      3rd Qu.:1.187e+07
## Max.      :1.348e+09      Max.      :1.264e+09      Max.      :1.154e+09
##
## X1980.Population      X1970.Population      Area..km..      Density..per.km..
## Min.      :      733      Min.      :      752      Min.      :      1      Min.      :      0.026
## 1st Qu.: 229614      1st Qu.: 155997      1st Qu.: 2650      1st Qu.: 38.418
## Median : 3141146      Median : 2604830      Median : 81200      Median : 95.347
## Mean      :18984617      Mean      :15786909      Mean      : 581449      Mean      : 452.127
## 3rd Qu.: 9826054      3rd Qu.: 8817329      3rd Qu.: 430426      3rd Qu.: 238.933
## Max.      :982372466      Max.      :822534450      Max.      :17098242      Max.      :23172.267
##
## Growth.Rate      World.Population.Percentage      Ranking.Category
## Min.      :0.912      Min.      : 0.0000      Length:234
## 1st Qu.:1.002      1st Qu.: 0.0100      Class :character
## Median :1.008      Median : 0.0700      Mode  :character
## Mean      :1.010      Mean      : 0.4271
## 3rd Qu.:1.017      3rd Qu.: 0.2800
## Max.      :1.069      Max.      :17.8800
```

```
head(mydata, n=10)
```

##	Rank	CCA3	Country	Capital	Continent	X2022.Population
## 1	1	CHN	China	Beijing	Asia	1425887337
## 2	2	IND	India	New Delhi	Asia	1417173173
## 3	3	USA	United States	Washington, D.C.	North America	338289857
## 4	4	IDN	Indonesia	Jakarta	Asia	275501339
## 5	5	PAK	Pakistan	Islamabad	Asia	235824862
## 6	6	NGA	Nigeria	Abuja	Africa	218541212
## 7	7	BRA	Brazil	Brasilia	South America	215313498
## 8	8	BGD	Bangladesh	Dhaka	Asia	171186372
## 9	9	RUS	Russia	Moscow	Europe	144713314
## 10	10	MEX	Mexico	Mexico City	North America	127504125
##	X2020.Population		X2015.Population	X2010.Population	X2000.Population	
## 1	1424929781		1393715448	1348191368	1264099069	
## 2	1396387127		1322866505	1240613620	1059633675	
## 3	335942003		324607776	311182845	282398554	
## 4	271857970		259091970	244016173	214072421	
## 5	227196741		210969298	194454498	154369924	
## 6	208327405		183995785	160952853	122851984	
## 7	213196304		205188205	196353492	175873720	
## 8	167420951		157830000	148391139	129193327	
## 9	145617329		144668389	143242599	146844839	
## 10	125998302		120149897	112532401	97873442	
##	X1990.Population		X1980.Population	X1970.Population	Area..km..	
## 1	1153704252		982372466	822534450	9706961	
## 2	870452165		696828385	557501301	3287590	
## 3	248083732		223140018	200328340	9372610	
## 4	182159874		148177096	115228394	1904569	
## 5	115414069		80624057	59290872	881912	
## 6	95214257		72951439	55569264	923768	
## 7	150706446		122288383	96369875	8515767	
## 8	107147651		83929765	67541860	147570	
## 9	148005704		138257420	130093010	17098242	
## 10	81720428		67705186	50289306	1964375	
##	Density..per.km..		Growth.Rate	World.Population.Percentage		
## 1	146.8933		1.0000	17.88		
## 2	431.0675		1.0068	17.77		
## 3	36.0935		1.0038	4.24		
## 4	144.6529		1.0064	3.45		
## 5	267.4018		1.0191	2.96		
## 6	236.5759		1.0241	2.74		
## 7	25.2841		1.0046	2.70		
## 8	1160.0350		1.0108	2.15		
## 9	8.4636		0.9973	1.81		
## 10	64.9082		1.0063	1.60		
##	Ranking.Category					
## 1	Very Highly Populated					
## 2	Very Highly Populated					
## 3	Neutral					
## 4	Very Highly Populated					
## 5	Very Highly Populated					
## 6	Highly Populated					
## 7	Less Populated					
## 8	Very Highly Populated					

```
## 9      Populated
## 10     Neutral
```

```
tail(mydata, n=10)
```

##	Rank	CCA3	Country	Capital	Continent
## 225	225	NRU	Nauru	Yaren	Oceania
## 226	226	WLF	Wallis and Futuna	Mata-Utu	Oceania
## 227	227	TUV	Tuvalu	Funafuti	Oceania
## 228	228	BLM	Saint Barthelemy	Gustavia	North America
## 229	229	SPM	Saint Pierre and Miquelon	Saint-Pierre	North America
## 230	230	MSR	Montserrat	Brades	North America
## 231	231	FLK	Falkland Islands	Stanley	South America
## 232	232	NIU	Niue	Alofi	Oceania
## 233	233	TKL	Tokelau	Nukunonu	Oceania
## 234	234	VAT	Vatican City	Vatican City	Europe
##	X2022.Population		X2020.Population	X2015.Population	X2010.Population
## 225	12668		12315	11185	10241
## 226	11572		11655	12182	13142
## 227	11312		11069	10877	10550
## 228	10967		10681	9643	8988
## 229	5862		5906	5978	6052
## 230	4390		4500	5059	4938
## 231	3780		3747	3408	3187
## 232	1934		1942	1847	1812
## 233	1871		1827	1454	1367
## 234	510		520	564	596
##	X2000.Population		X1990.Population	X1980.Population	X1970.Population
## 225	10377		9598	7635	6663
## 226	14723		13454	11315	9377
## 227	9638		9182	7731	5814
## 228	7082		5168	2983	2417
## 229	6274		6324	6106	5537
## 230	5138		10805	11452	11402
## 231	3080		2332	2240	2274
## 232	2074		2533	3637	5185
## 233	1666		1669	1647	1714
## 234	651		700	733	752
##	Area..km..	Density..per.km..	Growth.Rate	World.Population.Percentage	
## 225	21	603.2381	1.0125	0	
## 226	142	81.4930	0.9953	0	
## 227	26	435.0769	1.0096	0	
## 228	21	522.2381	1.0098	0	
## 229	242	24.2231	0.9964	0	
## 230	102	43.0392	0.9939	0	
## 231	12173	0.3105	1.0043	0	
## 232	260	7.4385	0.9985	0	
## 233	12	155.9167	1.0119	0	
## 234	1	510.0000	0.9980	0	
##	Ranking.Category				
## 225	Least Populated				
## 226	Least Populated				
## 227	Least Populated				
## 228	Neutral				
## 229	Neutral				
## 230	Neutral				
## 231	Less Populated				
## 232	Least Populated				

```
## 233 Least Populated
## 234      Populated
```

```
nrow(mydata)
```

```
## [1] 234
```

```
ncol(mydata)
```

```
## [1] 18
```

```
str(mydata)
```

```
## 'data.frame':    234 obs. of  18 variables:
## $ Rank           : int  1 2 3 4 5 6 7 8 9 10 ...
## $ CCA3           : chr  "CHN" "IND" "USA" "IDN" ...
## $ Country        : chr  "China" "India" "United States" "Indonesia" ...
## $ Capital        : chr  "Beijing" "New Delhi" "Washington, D.C." "Jakarta"
...
## $ Continent      : chr  "Asia" "Asia" "North America" "Asia" ...
## $ X2022.Population : int  1425887337 1417173173 338289857 275501339 235824862 2
18541212 215313498 171186372 144713314 127504125 ...
## $ X2020.Population : int  1424929781 1396387127 335942003 271857970 227196741 2
08327405 213196304 167420951 145617329 125998302 ...
## $ X2015.Population : int  1393715448 1322866505 324607776 259091970 210969298 1
83995785 205188205 157830000 144668389 120149897 ...
## $ X2010.Population : int  1348191368 1240613620 311182845 244016173 194454498 1
60952853 196353492 148391139 143242599 112532401 ...
## $ X2000.Population : int  1264099069 1059633675 282398554 214072421 154369924 1
22851984 175873720 129193327 146844839 97873442 ...
## $ X1990.Population : int  1153704252 870452165 248083732 182159874 115414069 95
214257 150706446 107147651 148005704 81720428 ...
## $ X1980.Population : int  982372466 696828385 223140018 148177096 80624057 7295
1439 122288383 83929765 138257420 67705186 ...
## $ X1970.Population : int  822534450 557501301 200328340 115228394 59290872 5556
9264 96369875 67541860 130093010 50289306 ...
## $ Area..km..     : int  9706961 3287590 9372610 1904569 881912 923768 8515767
147570 17098242 1964375 ...
## $ Density..per.km.. : num  146.9 431.1 36.1 144.7 267.4 ...
## $ Growth.Rate      : num  1 1.01 1 1.01 1.02 ...
## $ World.Population.Percentage: num  17.88 17.77 4.24 3.45 2.96 ...
## $ Ranking.Category  : chr  "Very Highly Populated" "Very Highly Populated" "Neut
ral" "Very Highly Populated" ...
```

#4 Transform at least one variable

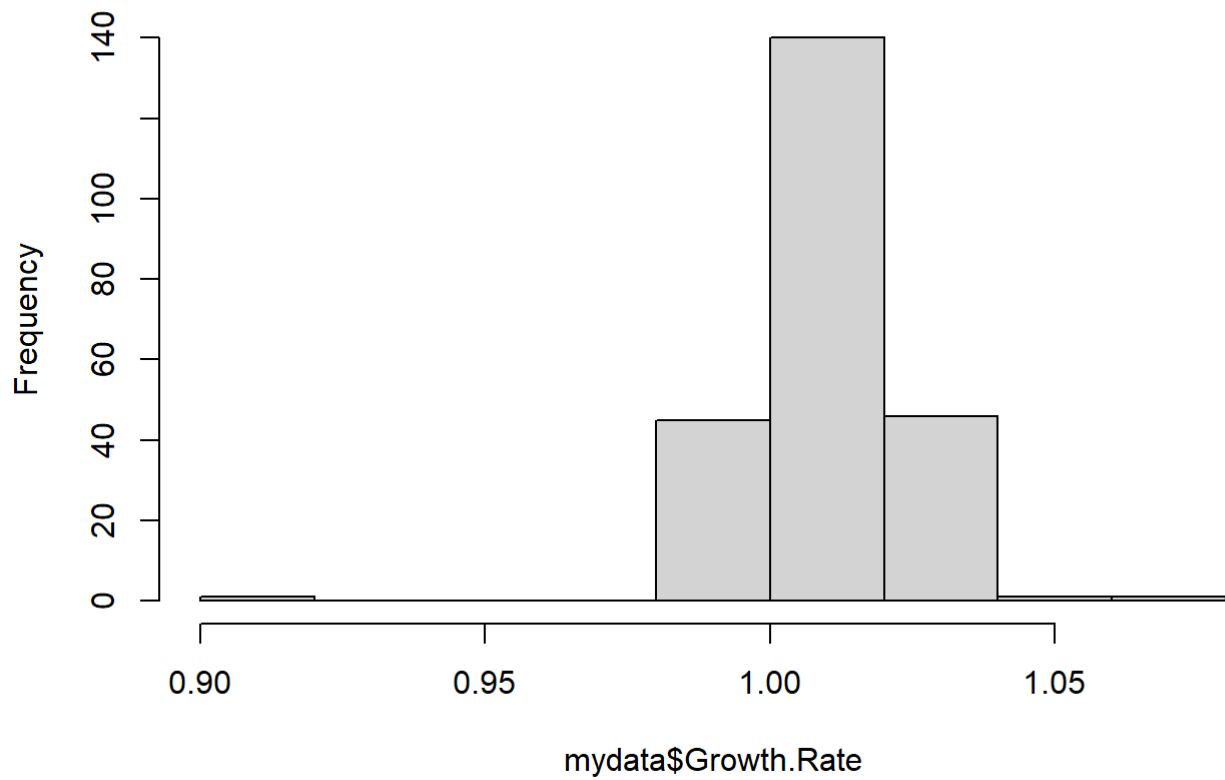
```
population_in_millions <- mydata$X2022.Population / 1000000
population_in_millions
```

##	[1]	1425.887337	1417.173173	338.289857	275.501339	235.824862	218.541212
##	[7]	215.313498	171.186372	144.713314	127.504125	123.951692	123.379924
##	[13]	115.559009	110.990103	99.010212	98.186856	88.550570	85.341241
##	[19]	83.369843	71.697030	67.508936	65.497748	64.626628	59.893885
##	[25]	59.037474	54.179306	54.027487	51.874024	51.815810	47.558630
##	[31]	47.249585	46.874204	45.510318	44.903225	44.496122	41.128771
##	[37]	39.857145	39.701739	38.454327	37.457971	36.408820	35.588987
##	[43]	34.627652	34.049588	33.938221	33.696614	33.475870	32.969517
##	[49]	30.547580	29.611714	28.301696	28.160542	27.914536	26.207977
##	[55]	26.177413	26.069416	23.893394	22.673762	22.593590	22.125249
##	[61]	21.832143	20.405317	20.017675	19.659267	19.603733	19.397998
##	[67]	18.001000	17.843908	17.723315	17.597511	17.564014	17.316449
##	[73]	16.767842	16.320537	13.859341	13.776698	13.352864	12.889576
##	[79]	12.356117	12.224110	11.655930	11.584996	11.285869	11.228821
##	[85]	11.212191	10.913164	10.549347	10.493986	10.432860	10.384971
##	[91]	10.358074	10.270865	10.142619	9.967308	9.952787	9.534954
##	[97]	9.441129	9.038309	8.939617	8.848699	8.740472	8.605718
##	[103]	7.529475	7.488865	7.221365	6.948392	6.812341	6.781953
##	[109]	6.780744	6.630623	6.430770	6.336392	5.975689	5.970424
##	[115]	5.882261	5.643453	5.579144	5.540745	5.489739	5.434319
##	[121]	5.302681	5.250072	5.185288	5.180829	5.023109	4.736139
##	[127]	4.576298	4.408581	4.268873	4.030358	3.744385	3.684032
##	[133]	3.422794	3.398366	3.272996	3.252407	3.233526	2.842321
##	[139]	2.827377	2.780469	2.750055	2.705992	2.695122	2.630296
##	[145]	2.567012	2.388992	2.305825	2.119844	2.105566	2.093599
##	[151]	1.850651	1.674908	1.531044	1.472233	1.341296	1.326062
##	[157]	1.299469	1.251488	1.201670	1.120849	0.974052	0.929766
##	[163]	0.836774	0.808726	0.782455	0.724273	0.695168	0.647599
##	[169]	0.627082	0.618040	0.593149	0.575986	0.533286	0.523787
##	[175]	0.449002	0.409984	0.405272	0.395752	0.372899	0.367507
##	[181]	0.326740	0.326101	0.306279	0.304557	0.289950	0.281635
##	[187]	0.227380	0.222382	0.191163	0.179857	0.171774	0.131232
##	[193]	0.125438	0.114164	0.110778	0.107118	0.106858	0.106445
##	[199]	0.103948	0.099465	0.093763	0.084519	0.079824	0.072737
##	[205]	0.068706	0.064184	0.063301	0.056466	0.053090	0.049551
##	[211]	0.047657	0.045703	0.044273	0.044175	0.041569	0.039327
##	[217]	0.036469	0.033660	0.032649	0.031791	0.031305	0.018055
##	[223]	0.017011	0.015857	0.012668	0.011572	0.011312	0.010967
##	[229]	0.005862	0.004390	0.003780	0.001934	0.001871	0.000510

#5.1 Plot at least one quantitative variable

```
hist(mydata$Growth.Rate)
```

Histogram of mydata\$Growth.Rate



#5.2 Scatter plot

```
x=mydata$X2022.Population  
y=mydata$Rank  
plot(x,y)
```

