

The dataset *ToyotaCorolla.csv* contains data on used cars on sale during the late summer of 2004 in the Netherlands. It has 1436 records containing details on 38 attributes, including *Price*, *Age*, *Kilometers*, *HP*, and other specifications.

Import the “Universities2019.csv” file to R Studio and write R code to produce summary statistics of all numeric variables. For each numeric variable, include the following.

Variable name
Minimum
Maximum
Median
Mean
Standard Deviation
Number of Missing Values

Paste a screenshot of your summary statistics below.

```
> university_df <- df[,c(4:8)]
> View(university_df)
> summary_statistics <- data.frame( mean = sapply(university_df, mean, na.rm = TRUE),
+                                   median = sapply(university_df, median, na.rm = TRUE),
+                                   sd = sapply(university_df, sd, na.rm = TRUE),
+                                   variance = sapply(university_df, var, na.rm = TRUE),
+                                   min = sapply(university_df, min, na.rm = TRUE),
+                                   max = sapply(university_df, max, na.rm = TRUE),
+                                   count = sapply(university_df, length),
+                                   miss.val = sapply(university_df, function(x)
+                                     sum(length(which(is.na(x)))))
+ )
> print(summary_statistics)
```

	mean	median	sd	variance	min	max	count	miss.val
Applicants	2752.09752	1470.0	3541.974712	1.254558e+07	35.0	48094.0	1302	10
Admits	1870.68319	1095.0	2250.866400	5.066400e+06	35.0	26330.0	1302	11
Enrolled	778.88049	447.0	884.578274	7.824787e+05	18.0	7425.0	1302	5
StudentFacultyRatio	14.85877	14.3	5.186399	2.689873e+01	2.3	91.8	1302	2
GraduationRate	60.40532	60.0	18.889058	3.567965e+02	8.0	118.0	1302	98

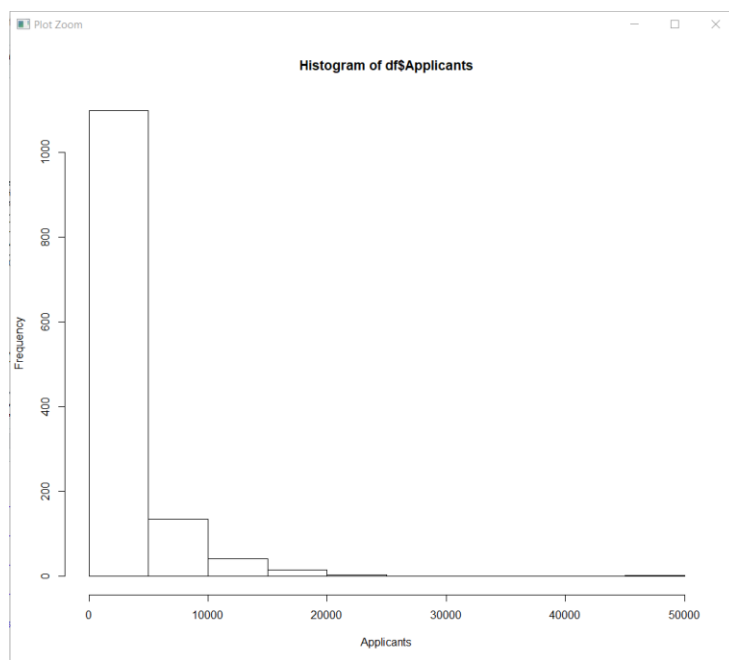
```
> View(summary_statistics)
```

	mean	median	sd	variance	min	max	count	miss.val
Applicants	2752.09752	1470.0	3541.974712	1.254558e+07	35.0	48094.0	1302	10
Admits	1870.68319	1095.0	2250.866400	5.066400e+06	35.0	26330.0	1302	11
Enrolled	778.88049	447.0	884.578274	7.824787e+05	18.0	7425.0	1302	5
StudentFacultyRatio	14.85877	14.3	5.186399	2.689873e+01	2.3	91.8	1302	2
GraduationRate	60.40532	60.0	18.889058	3.567965e+02	8.0	118.0	1302	98

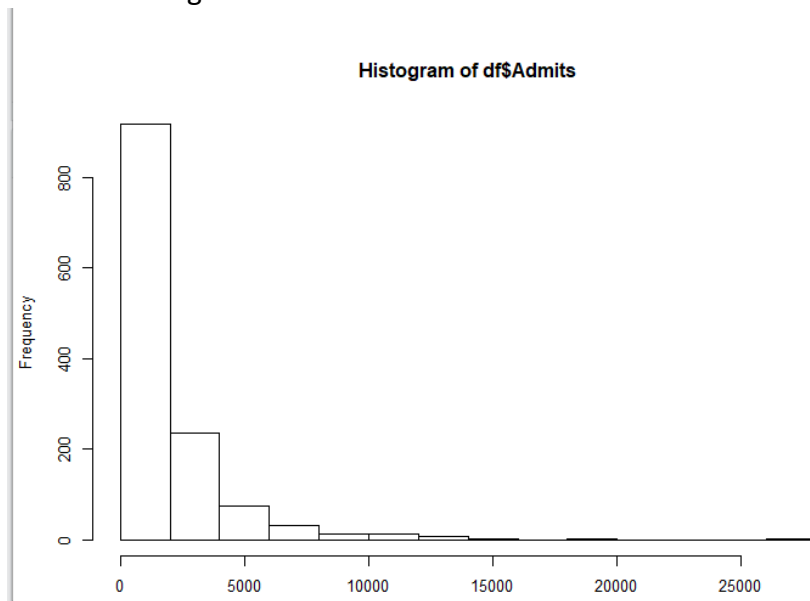
Create a new column called “AdmitRate” and add it to the data set as the last column. Admit rate should be calculated by dividing “Admits” by “Applicants” to get the percent of students admitted. Paste a screen shot of the first 10 rows of the dataset with the new column below.

	Name	State	Type	Applicants	Admits	Enrolled	StudentFacultyRatio	GraduationRate	AdmitRate
1	Abilene Christian University	TX	Private	1660	1232	721	18.1	60	0.7421687
2	Adams State College	CO	Public	1508	1259	569	27.9	60	0.8348806
3	Adelphi University	NY	Private	2186	1924	512	12.2	56	0.8801464
4	Adrian College	MI	Private	1428	1097	336	12.9	54	0.7682073
5	Agnes Scott College	GA	Private	417	349	137	7.7	59	0.8369305
6	Alabama Agri. & Mech. Univ.	AL	Public	2817	1920	984	14.3	40	0.6815761
7	Alabama State University	AL	Public	4639	3272	1278	18.7	15	0.7053244
8	Alaska Pacific University	AK	Private	193	146	55	11.9	15	0.7564767
9	Albany State College	GA	Public	1461	1097	580	17.0	62	0.7508556
10	Albertson College	ID	Private	587	479	158	9.4	55	0.8160136

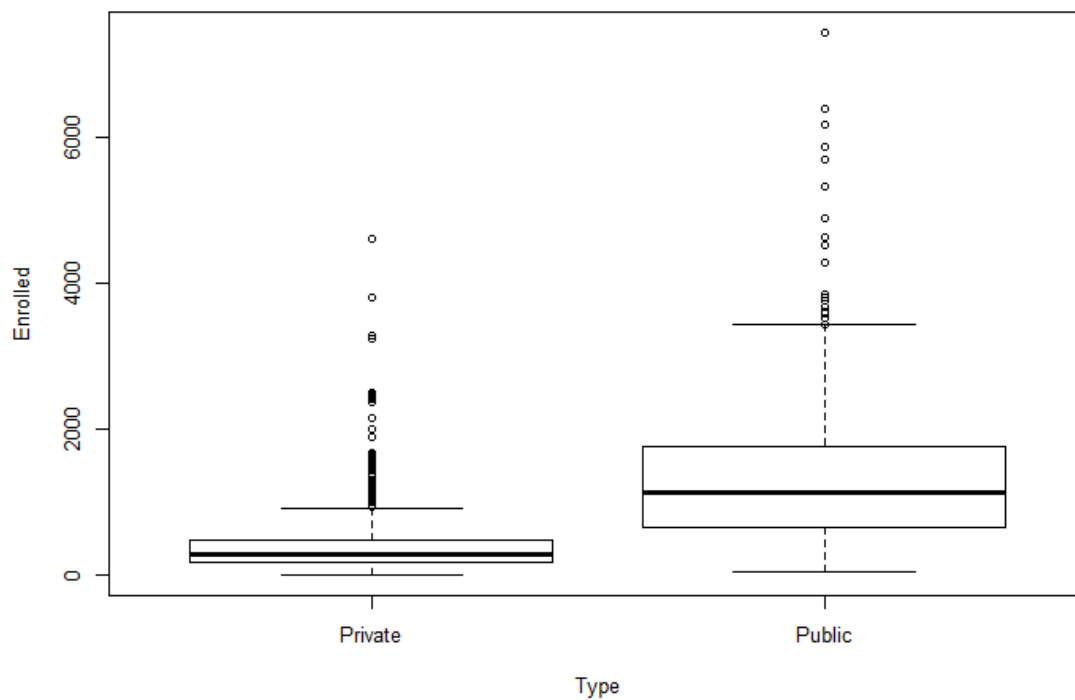
Create a histogram of the Applicants variable.



Create a histogram of Admits.



Create a side by side Box Plot for Enrolled using Type as the by variable.



Create a scatter plot of AdmitRate and GraduationRate

