MM 217: Plotting and Pie Charts in R

In the previous tutorials, you have learnt some basics of R as well as some of its plotting capabilities. We continue with the descriptive statistics aspects and learn the same using R. Specifically, in this tutorial, I will show a few scripts and the figures generated by them. I want you to study these scripts and the associated plots and understand what each command in the script is achieving – on your own, using help commands as and when necessary. Based on the understanding, you will solve a few problems of your own on a new data set.

For our practice exercise, we will be using the following data set which gives the data on the areas and populations of the states and union territories in India. The data is taken from the website of Ministry of Statistics and Programme Implementation, Government of India : http://mospi.nic.in/statistical-year-book-india/2018

State UT	Area	Males	Females	Persons	Rural	Urban
Andhra Pradesh	275045	42442146	42138631	84580777	56361702	28219075
Arunachal Pradesh	83743	713912	669815	1383727	1066358	317369
Assam	78438	15939443	15266133	31205576	26807034	4398542
Bihar	94163	54278157	49821295	104099452	92341436	11758016
Chhattisgarh	135192	12832895	12712303	25545198	19607961	5937237
Goa	3702	739140	719405	1458545	551731	906814
Gujarat	196244	31491260	28948432	60439692	34694609	25745083
Haryana	44212	13494734	11856728	25351462	16509359	8842103
Himachal Pradesh	55673	3481873	3382729	6864602	6176050	688552
Jammu & Kashmir	222236	6640662	5900640	12541302	9108060	3433242
Jharkhand	79716	16930315	16057819	32988134	25055073	7933061
Karnataka	191791	30966657	30128640	61095297	37469335	23625962
Kerala	38852	16027412	17378649	33406061	17471135	15934926
Madhya Pradesh	308252	37612306	35014503	72626809	52557404	20069405
Maharashtra	307713	58243056	54131277	112374333	61556074	50818259
Manipur	22327	1438586	1417208	2855794	2021640	834154
Meghalaya	22429	1491832	1475057	2966889	2371439	595450
Mizoram	21081	555339	541867	1097206	525435	571771
Nagaland	16579	1024649	953853	1978502	1407536	570966
Odisha	155707	21212136	20762082	41974218	34970562	7003656
Punjab	50362	14639465	13103873	27743338	17344192	10399146
Rajasthan	342239	35550997	32997440	68548437	51500352	17048085
Sikkim	7096	323070	287507	610577	456999	153578
Tamil Nadu	130060	36137975	36009055	72147030	37229590	34917440
Tripura	10486	1874376	1799541	3673917	2712464	961453
Uttar Pradesh	240928	104480510	95331831	199812341	155317278	44495063
Uttarakhand	53483	5137773	4948519	10086292	7036954	3049338
West Bengal	88752	46809027	44467088	91276115	62183113	29093002
A.& N.Islands	8249	202871	177710	380581	237093	143488
Chandigarh	114	580663	474787	1055450	28991	1026459
D.& N.Haveli	491	193760	149949	343709	183114	160595
Daman & Diu	111	150301	92946	243247	60396	182851
Delhi	1483	8987326	7800615	16787941	419042	16368899
Lakshadweep	30	33123	31350	64473	14141	50332
Puducherry	490	612511	635442	1247953	395200	852753

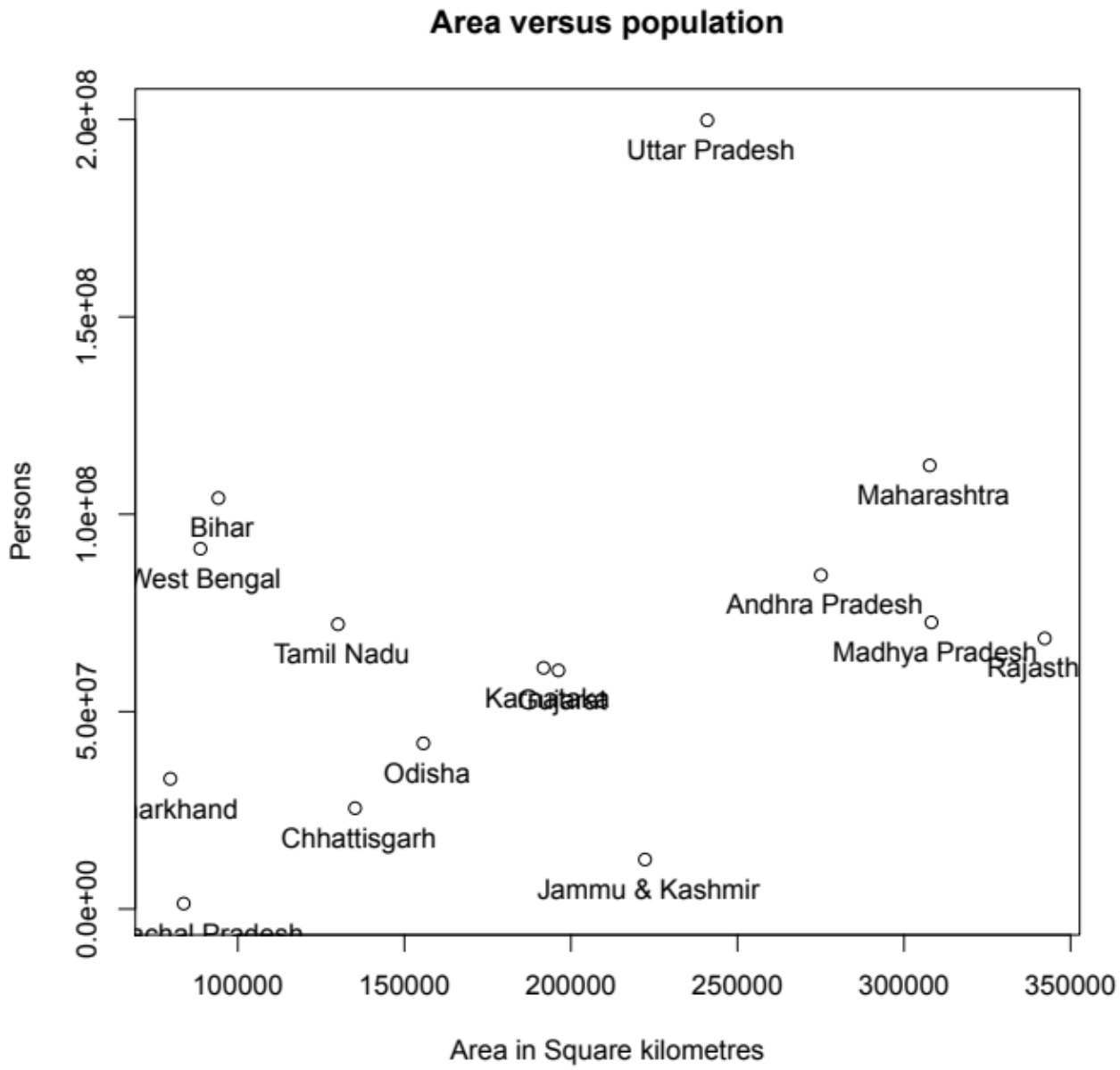
This data file is uploaded with the name AreaPopulationData.csv.

Given this data, now, let us consider the following scripts:

Script 1:

```
X <- read.csv("AreaPopulationData.csv")
ordered.states <- head(X[order(X$Area,decreasing=TRUE),],15)
str(ordered.states)
plot(Persons~Area,data=ordered.states,main="Area versus population",xlab="Area in Square
kilometres",ylab="Persons")
text(Persons~Area,data=ordered.states,labels=(ordered.states[1:15,1]),adj=c(0.5,2))
pdf("T3S1.pdf")
plot(Persons~Area,data=ordered.states,main="Area versus population",xlab="Area in Square
kilometres",ylab="Persons")
text(Persons~Area,data=ordered.states,labels=(ordered.states[1:15,1]),adj=c(0.5,2))
dev.off
```

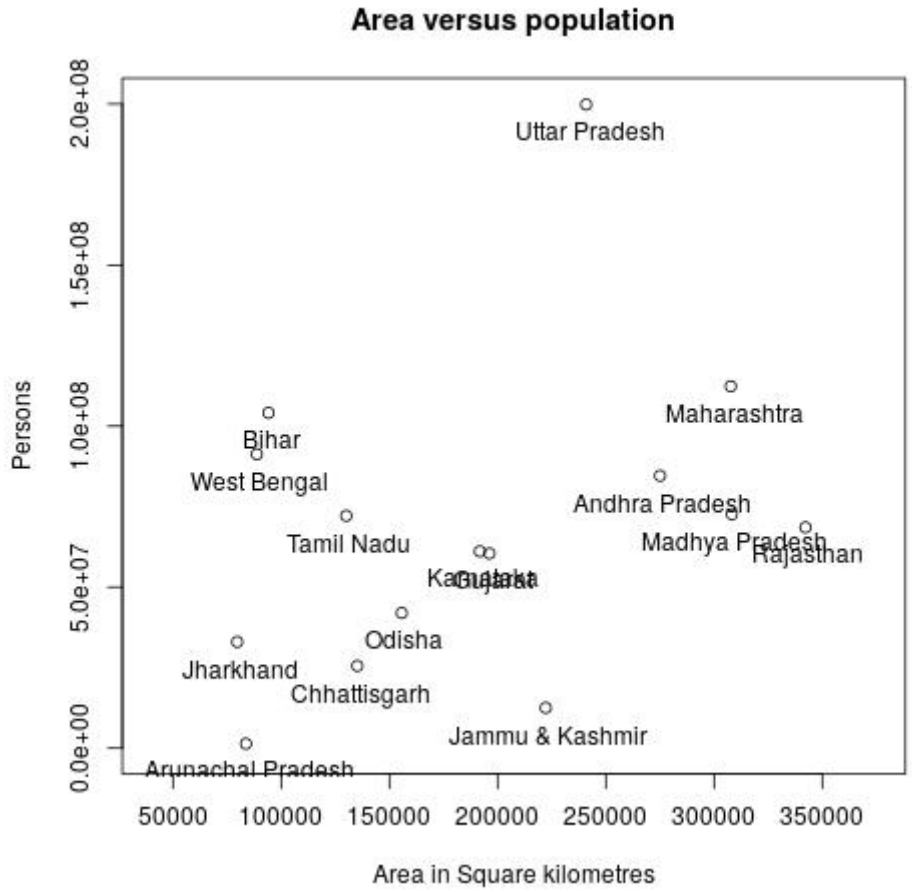
The figure generated by this script is as shown below.



Script 2:

```
X <- read.csv("AreaPopulationData.csv")
ordered.states <- head(X[order(X$Area,decreasing=TRUE),],15)
str(ordered.states)
jpeg("T3S2.jpg")
plot(Persons~Area,data=ordered.states,main="Area versus population",xlab="Area in Square
kilometres",ylab="Persons",xlim = c(40000,375000),ylim=c(0,2.0e8))
text(Persons~Area,data=ordered.states,labels=(ordered.states[1:15,1]),adj=c(0.5,2))
dev.off
```

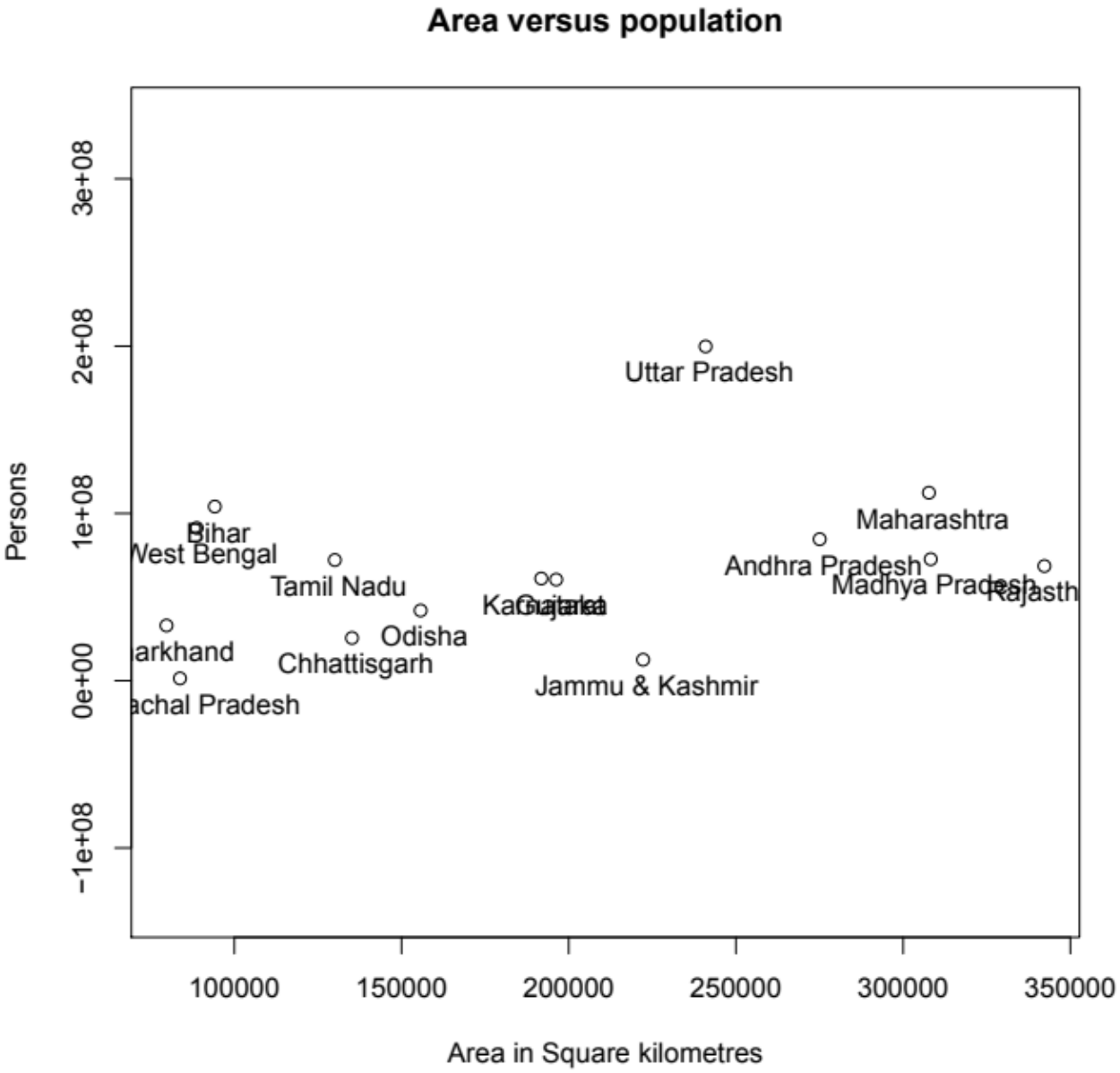
The figure generated by this script is as shown below.



Script 3:

```
X <- read.csv("AreaPopulationData.csv")
ordered.states <- head(X[order(X$Area,decreasing=TRUE),],15)
str(ordered.states)
pdf("T3S3.pdf")
plot(Persons~Area,data=ordered.states,main="Area versus population",xlab="Area in Square
kilometres",ylab="Persons",asp="0.0005")
text(Persons~Area,data=ordered.states,labels=(ordered.states[1:15,1]),adj=c(0.5,1.8))
dev.off
```

The figure generated by this script is as shown below.



Script 4

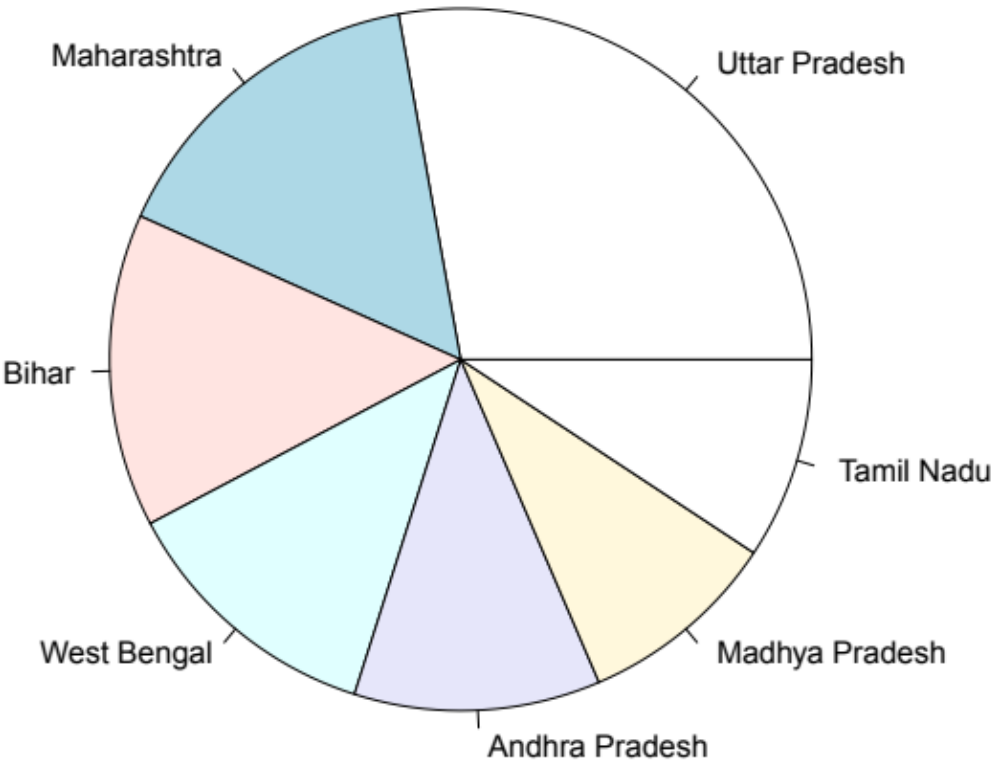
```
X <- read.csv("AreaPopulationData.csv")
ordered.states <- head(X[order(X$Persons,decreasing=TRUE),],7)
str(ordered.states)
z = ordered.states[,5]
z = as.integer(100*z/(sum(z)))
pie(z,labels=ordered.states[1:7,1])
pdf("PopPieChart1.pdf")
pie(z,labels=ordered.states[1:7,1])
dev.off
```

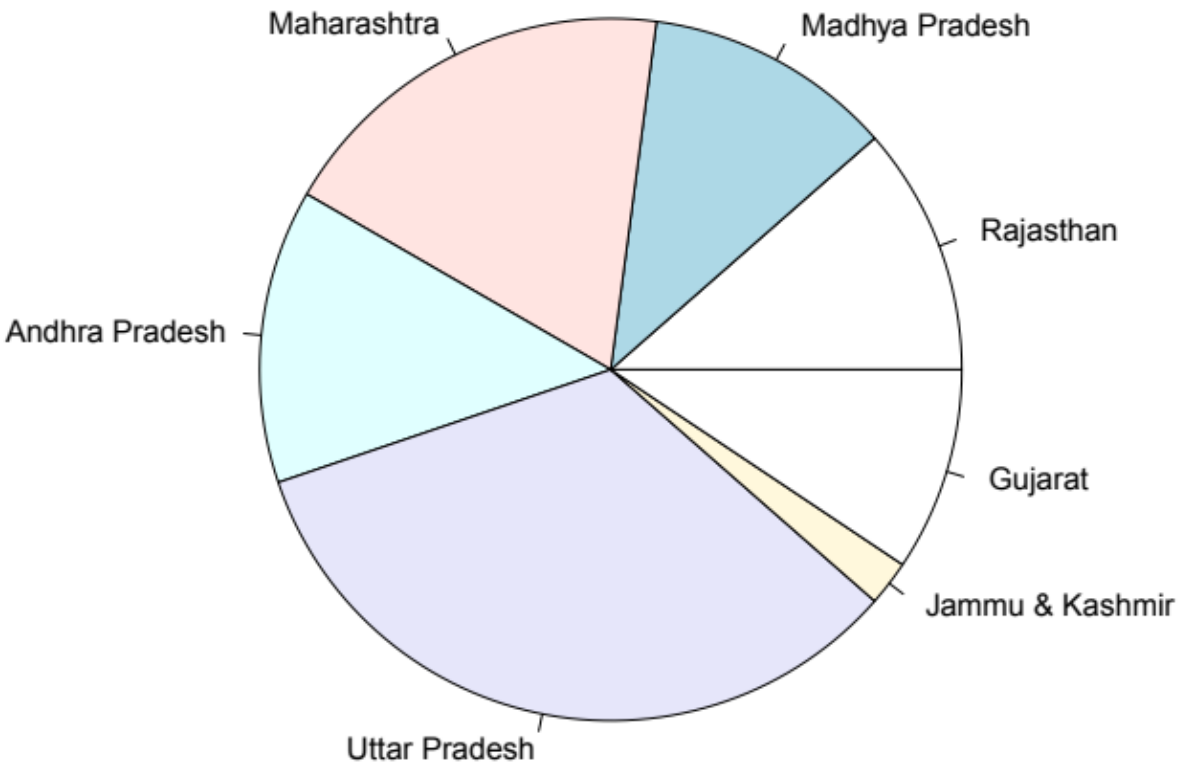
Script 5

```
X <- read.csv("AreaPopulationData.csv")
```

```
ordered.states <- head(X[order(X$Area,decreasing=TRUE),],7)
str(ordered.states)
z = ordered.states[,5]
z = as.integer(100*z/(sum(z)))
pie(z,labels=ordered.states[1:7,1])
pdf("PopPieChart2.pdf")
pie(z,labels=ordered.states[1:7,1])
dev.off
```

The figures generated by these two scripts are as shown below.





Based on your understanding of the scripts, you can solve new plotting problems. As an exercise, let us consider the following data – again taken from the website of Ministry of Statistics and Programme Implementation, Government of India :
http://mospi.nic.in/statistical-year-book-india/2018

This data is given to you as the csv file FactorySurvey.csv

The amounts are all in tens of millions of rupees. The rest are numbers.

State UT	Number of Factories	FixedCapital	Working Capital	Productive Capital	Invested Capital	Number of workers
Andhra Pradesh	16012	148244	5681	153926	195906	424075
Arunachal Pradesh	124	192	360	552	430	2925
Assam	3717	16169	11138	27307	25829	163348
Bihar	3529	9945	3492	13436	15897	125557
Chhattisgarh	2809	77935	70223	148158	97096	142799
Goa	635	8871	6227	15098	15185	45052
Gujarat	23433	437702	63253	500954	606559	1103204
Haryana	8243	87948	24786	112734	135903	580933
Himachal Pradesh	2784	48755	16403	65158	63451	147796
Jammu & Kashmir	965	5779	3185	8964	9061	52037
Jharkhand	2738	75665	1199	76865	103096	143595
Karnataka	12566	169023	45737	214761	233622	748372
Kerala	7320	29350	8524	37873	43655	294325
Madhya Pradesh	4240	97297	12931	110227	123602	256924
Maharashtra	28601	337144	109971	447114	497980	1305350
Manipur	160	114	169	283	179	6568
Meghalaya	109	3503	962	4465	4312	11870
Nagaland	197	200	167	367	310	4843
Odisha	2803	229479	-2979	226500	264952	214836
Punjab	12413	41140	32405	73545	83488	467951
Rajasthan	8986	76544	28456	105000	108239	375780
Sikkim	67	1762	1495	3257	2725	11100
Tamil Nadu	37878	221316	67133	288449	329168	1741427
Telangana	14427	57393	27979	85372	87598	585456
Tripura	548	324	22	346	499	26256
UttaraKhand	2987	51042	13174	64216	72122	295217
Uttar Pradesh	14867	108968	24529	133498	175273	673431
West Bengal	9112	86268	22413	108681	132137	504148
A & N. Island	20	23	-41	-18	38	427
Chandigarh	276	826	421	1247	1358	6534
Dadra & N Haveli	1411	23866	18776	42643	42405	96987
Daman & Diu	1874	9491	8064	17555	14319	78347
Delhi	3868	7138	12234	19372	19704	80766
Puducherry	716	5037	2351	7388	7870	37052

Problem 1

Plot the scatter plot of number of factories versus number of workers. Label the x-axis “Number of Factories” and the y-axis “Number of Workers”. Call the plot “Factories versus workers”. Against each data point, the name of the state should be mentioned. Plot only the top 10 states in terms of the number of workers.

Problem 2

Plot the scatter plot of number of factories versus number of workers. Label the x-axis “Number of Factories” and the y-axis “Number of Workers”. Call the plot “Factories versus workers”. Against each data point, the name of the state should be mentioned. Plot only the bottom 10 states in terms of the number of factories.

Problem 3

Plot the pie chart of invested captial of the top 10 states; make sure that the states are labelled.

Problem 4

Plot the pie chart of invested captial of the bottom 10 states; make sure that the states are labelled.

Problem 5

Write an R script to calculate the mean, median and standard deviation of invested capital. In terms of these quantities, can you comment on how Bihar, Madhyapradesh, Rajasthan and Uttarpradesh perform? Again, in terms of these quantities, can you comment on how Maharashtra, Gujarat, Tamilnadu and Andhrapradesh perform?

Problem 6

Write an R script to give an histogram of Working capital.

Problem 7

Write an R script to give a bar chart of Productive Capital versus Number of Workers.

Problem 8

Write an R script to calculate the correlation coefficients between (a) Number of factories and number of workers, (b) Number of workers and productive capital, and, (c) Working capital and number of factories. Comment on the three coefficients and indicate which has the strongest correlation. Why?