INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

Lab Report

Course: Data Analytics in R (CS6E23L)

Course Instructor: Dr. Kavi Mahesh

By:

Paka Sravan Kumar Yadav

6th Semester

3rd Year

16CS11

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

Lab – 01 (06th Feb 2019)

Year 2011:

Table VITAL EVENTS REGISTERED FROM 1971 TO 2011

⇒ a=read.csv(file = "G:\\Required\\6th Sem\\DA\\Lab\\Lab1\\2011.csv",head=TRUE, sep = ",", nrows = 41)

⇒ a

```
> a
   Year Live_births Still_births Deaths Vital_Births Vital_Deaths Percentage_Births Percentage_Deaths
1  1971      469226         9966 176160        16.00         6.00             50.50             40.60
2  1972      484616        13312 179593        16.10         6.00             51.10             47.20
3  1973      463130        11970 193725        14.94         6.24             51.70             50.30
4  1974      435353        10612 166102        13.74         5.24             49.10             48.10
5  1975      453444        10689 171857        13.97         5.29             50.40             48.10
6  1976      454851        10713 176061        13.68         5.30             46.40             45.30
7  1977      459473         9828 175824        12.92         5.17             48.20             46.20
8  1978      427336         4991 155489        12.22         4.44             42.00             37.30
9  1979      455668         7714 163296        12.82         4.59             44.70             45.00
10 1980      460295         7673  87556        12.68         2.41             45.90             25.10
11 1981      466387         7036 150526        12.56         4.06             44.40             44.60
12 1982      480337         4420 150008        12.66         3.96             45.40             43.00
13 1983      406812         5921 224115        10.51         3.21             36.10             34.50
14 1984      439892         6472 132568        11.14         3.35             36.80             35.00
15 1985      484334         2373 154186        12.03         3.83             40.60             43.50
16 1986      564500         5739 162700        13.65         3.93             47.40             45.20
17 1987      564015         4897 154018        13.40         3.66             46.40             42.10
18 1988      641846         5276 195787        14.98         4.57             52.20             51.90
19 1989      673287         4734 199766        17.68         5.09             63.10             57.80
20 1990      780496         5736 209873        17.61         4.73             62.90             58.40
21 1991      792291         4938 240206        17.65         5.35             65.60             59.40
22 1992      827188         5551 215666        18.00         4.69             68.50             55.20
23 1993      860471         3949 257133        18.34         5.48             71.90             68.50
24 1994      886320         5236 282180        21.60         6.90             86.70             83.10
25 1995      996077         5656 325279        20.34         6.64             84.40             87.40
26 1996     1028112         5415 336535        20.54         6.61             89.34             86.92
27 1997     1031329         4545 350264        20.17         6.85             88.86             90.14
28 1998     1042256         4213 374400        20.00         7.16             90.67             90.71
29 1999      997649         4473 346451        18.68         6.48             83.78             84.26
30 2000     1009716         5472 351736        20.04         6.98             91.07             89.48
31 2001     1017224         4557 365181        19.51         7.00             87.88             92.15
32 2002      973653         4187 355662        18.85         6.89             85.29             95.69
33 2003     1001749         3628 359661        19.31         6.93             88.58             96.25
34 2004      988520         3943 343644        18.82         6.54             90.05             94.78
35 2005     1007868         4538 364415        18.51         6.69             89.85             94.23
36 2006     1046531         5091 387604        18.95         7.02             94.28             98.87
37 2007     1046424         5526 381890        18.95         6.92             95.23             94.79
38 2008     1082450         5069 372062        19.30         6.63             97.47             89.59
39 2009     1076383         5729 373290        19.05         6.61             97.69             91.81
40 2010     1071518         6587 381743        18.29         6.51             95.26             91.69
41 2011     1108562         6940 384745        18.72         6.50             99.47             91.55
```

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

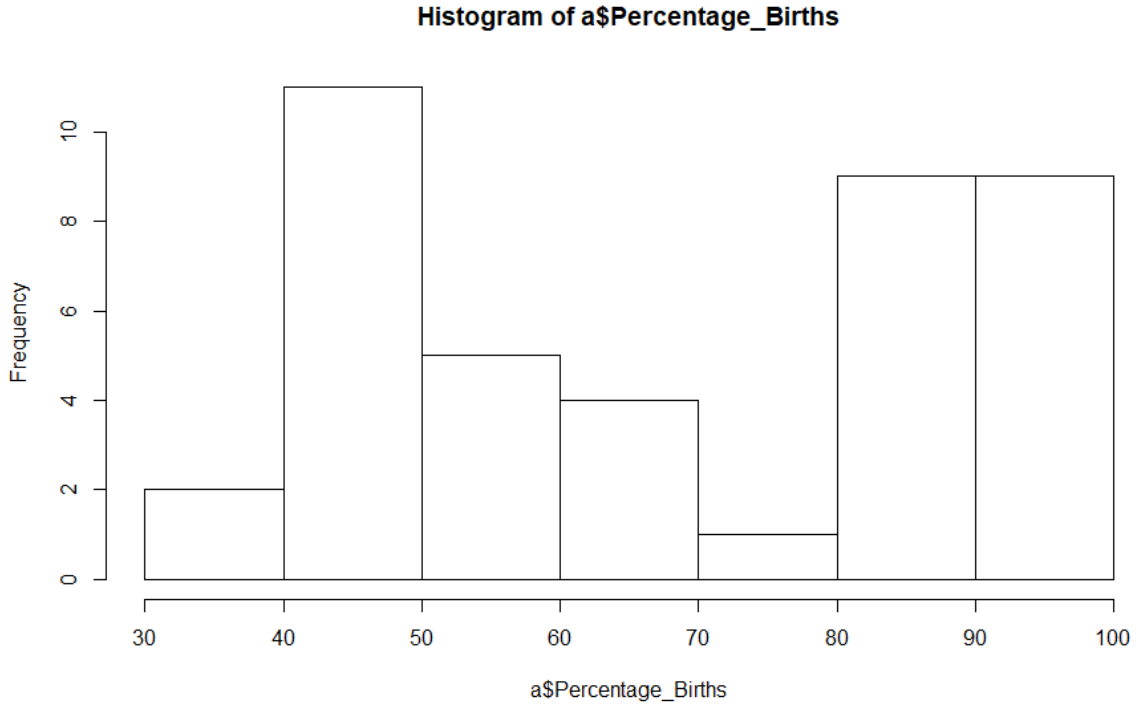
Basic Statistics:

⇒ mean(a$Live_births)
⇒ median(a$Still_births)
⇒ mode(a$Vital_Births)
⇒ var(a$Deaths)
⇒ sd(a$Percentage_Births)
⇒ IQR(a$Vital_Births)

```
> mean(a$Live_births)
[1] 755063.1
> median(a$Still_births)
[1] 5472
> mode(a$Vital_Births)
[1] "numeric"
> var(a$Deaths)
[1] 9175683900
> sd(a$Percentage_Births)
[1] 21.74335
> IQR(a$Vital_Births)
[1] 5.3
```

Histogram:

⇒ hist(a$Percentage_Births)



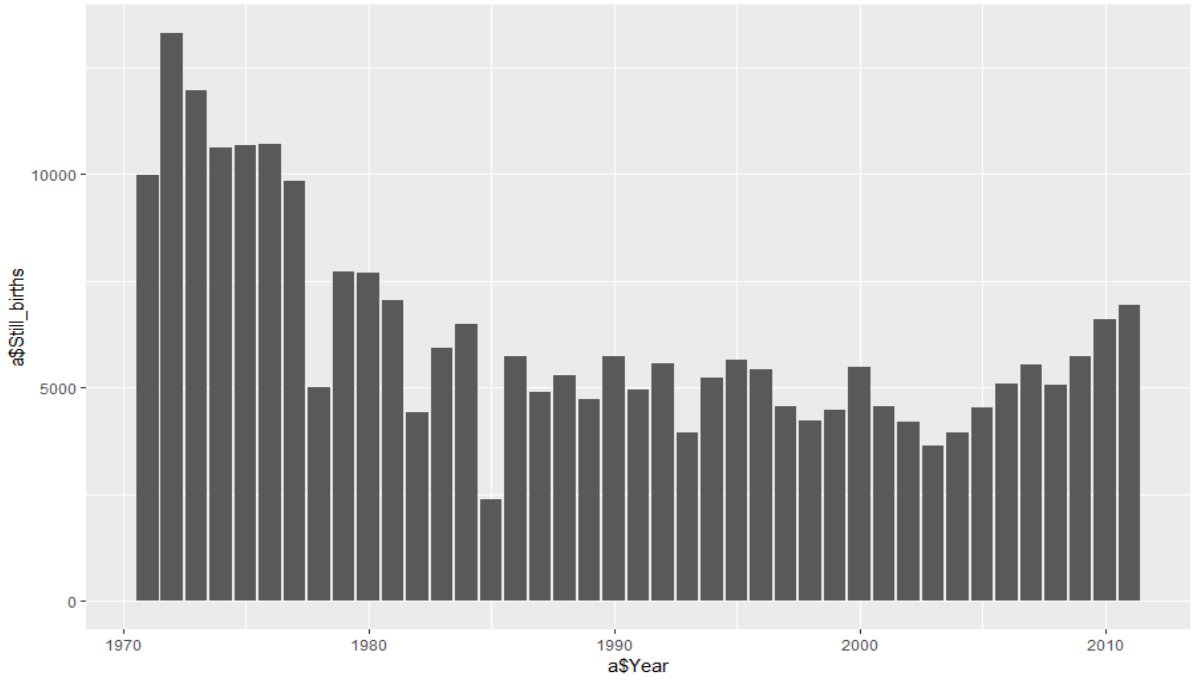
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

Line:

⇒ ggplot(a,aes(x=a$Year,y=a$Deaths))+ geom_line() + xlim(1970, 2011) + scale_y_continuous(limits=c(87000,390000))

BarGraph:

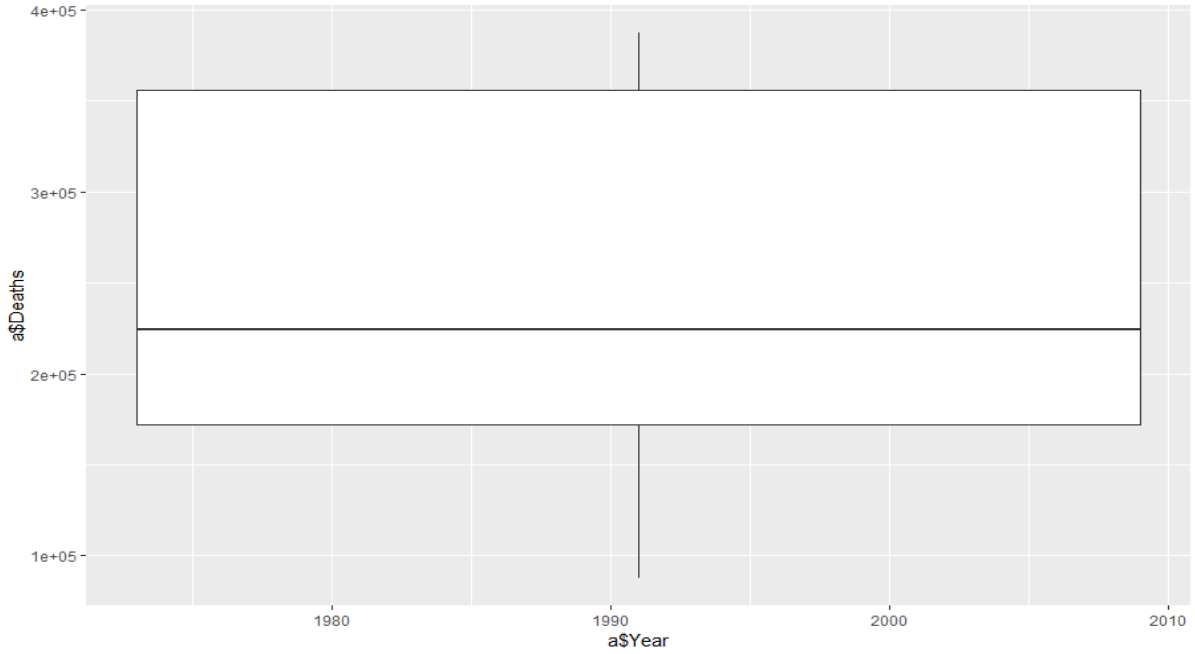
⇒ gplot(a, aes(x=a$Year, y=a$Still_births)) + geom_col()



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

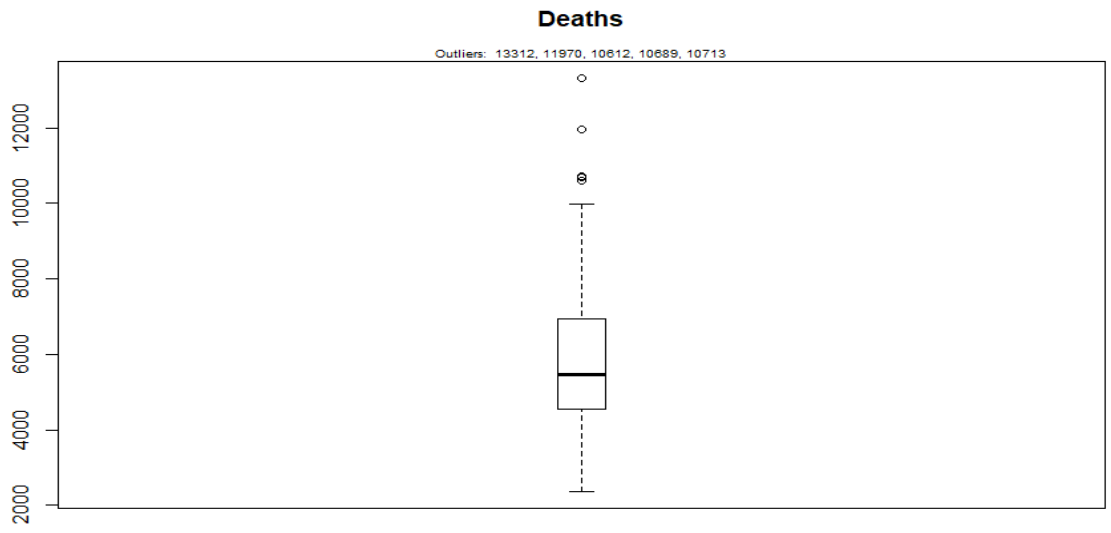
Boxplot:

⇒ ggplot(a, aes(x=a$Year,y=a$Deaths)) + geom_boxplot()



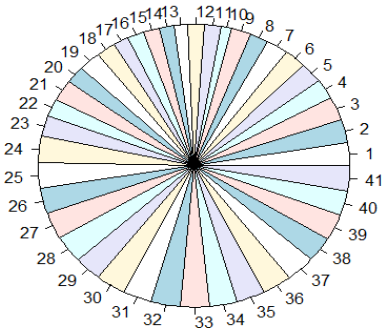
Outliers:

⇒ outlier=boxplot.stats(a$Still_births)$out
⇒ boxplot(a$Still_births, main="Deaths", boxwex=0.1)
⇒ mtext(paste("Outliers: ", paste(outlier, collapse = ", ")),cex=0.6)



PieChart:

⇒ pie(a$Vital_Deaths)



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

Table DISTRICTWISE REGISTERED BIRTHS, DEATHS, INFANT DEATHS, STILL BIRTHS AND RATES URBAN-2011:

⇒ b=read.csv(file = "G:\\Required\\6th Sem\\DA\\Lab\\Lab1\\2011.csv",head=TRUE, sep = ",", skip=46, nrows=30)

⇒ b

	Sl.No	District	Births_Registered	Birth_Rate	Death_Registered	Rate	Registered_Infant_Death	Still.Birth_Registered	Still.Birth_Rate
1	1	BAGALKOTE	35199	18.74	5119	6.06	234	426	10.35
2	2	BANGALORE (R)	7479	12.80	1722	5.66	27	4	0.29
3	3	BANGALORE (U)	141434	14.28	48611	4.88	1216	264	2.00
4	4	BELGAUM	54881	14.99	10951	6.80	516	868	13.37
5	5	BELLARY	28309	21.87	6728	6.00	372	335	6.85
6	6	BIDAR	24793	19.08	2370	3.98	188	343	4.18
7	7	BIJAPUR	29234	24.59	4040	6.47	398	327	4.35
8	8	CHAMARAJANAGAR	7070	18.18	1329	6.77	18	14	1.87
9	9	CHICKBALLAPUR	14337	18.50	1655	5.68	36	85	4.91
10	10	CHIKMAGALUR	14575	16.13	2091	6.51	91	38	2.06
11	11	CHITRADURGA	17439	16.20	3096	7.15	145	171	10.34
12	12	DAKSHINA KANNADA	30225	16.64	8758	5.05	514	333	8.40
13	13	DAVANAGERE	27982	19.04	7100	7.96	483	231	7.06
14	14	DHARWAD	34689	20.91	10170	6.84	1102	761	9.42
15	15	GADAG	16325	17.45	4104	8.37	170	309	11.27
16	16	GULBARGA	17430	22.41	4339	5.64	310	10	0.84
17	17	HASSAN	21600	16.15	3240	5.39	125	26	0.98
18	18	HAVERI	18665	20.99	2737	7.00	86	314	10.11
19	19	KODAGU	6146	14.05	884	6.47	114	23	2.63
20	20	KOLAR	17275	19.01	3508	5.75	81	217	9.09
21	21	KOPPAL	10513	24.31	1661	6.10	36	14	2.80
22	22	MANDYA	17392	16.68	2940	5.97	115	15	0.19
23	23	MYSORE	44644	16.77	11354	6.40	881	234	2.75
24	24	RAICHUR	18559	21.62	4114	6.65	249	277	9.27
25	25	RAMANAGAR	9712	22.55	1885	6.25	9	6	0.00
26	26	SHIMOGA	29433	16.48	5713	6.97	327	329	6.89
27	27	TUMKUR	29496	19.57	4428	6.60	110	32	1.14
28	28	UDUPI	16463	17.51	3770	5.12	125	136	0.67
29	29	UTTARA KANNADA	22150	17.69	3096	6.35	101		
30	30	YADGIR	4716	17.37	1074	5.20	19		

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

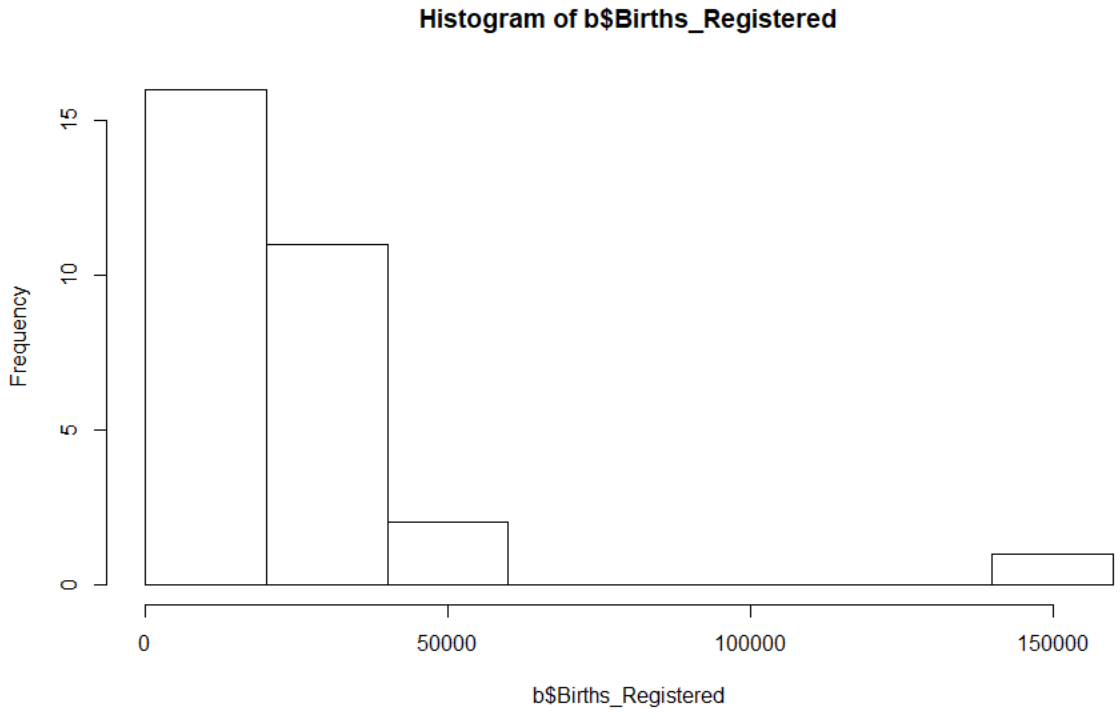
Basic Statistics:

⇒ mean(b$Births_Registered)
⇒ median(b$Birth_Rate)
⇒ mode(b$Death_Registered)
⇒ var(b$Rate)
⇒ sd(b$Registered_Infant_Death)
⇒ IQR(b$Still.Birth_Registered)
⇒ var(b$Still.Birth_Rate)

```
> mean(b$Births_Registered)
[1] 25605.5
> median(b$Birth_Rate)
[1] 17.935
> mode(b$Death_Registered)
[1] "numeric"
> var(b$Rate)
[1] 0.8180051
> sd(b$Registered_Infant_Death)
[1] 310.7645
> IQR(b$Still.Birth_Registered)
[1] 300
> var(b$Still.Birth_Rate)
[1] 16.31401
```

Histogram:

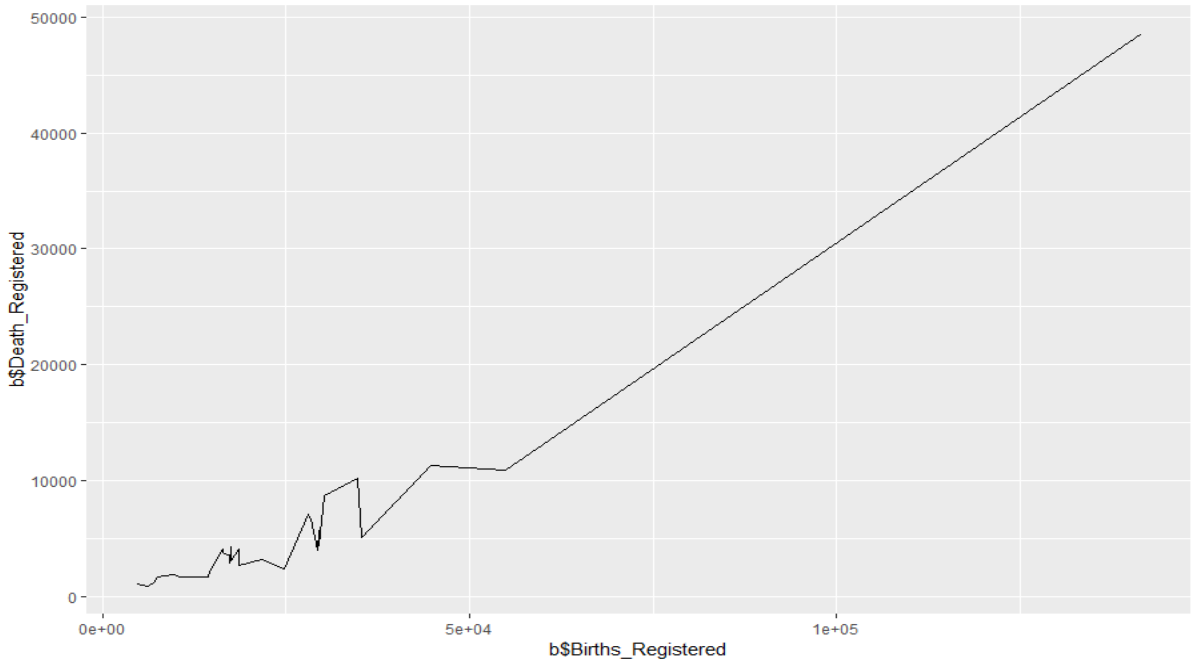
⇒ hist(b$Births_Registered)



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

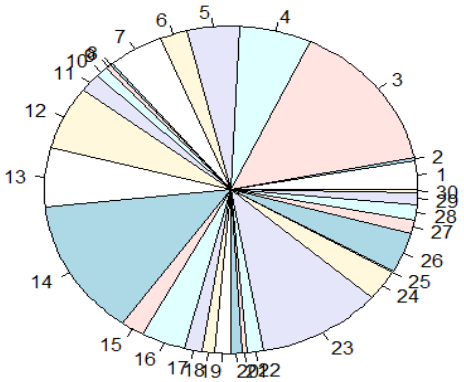
Line:

⇒ ggplot(b, aes(x=b$Births_Registered,y=b$Death_Registered)) + geom_line()



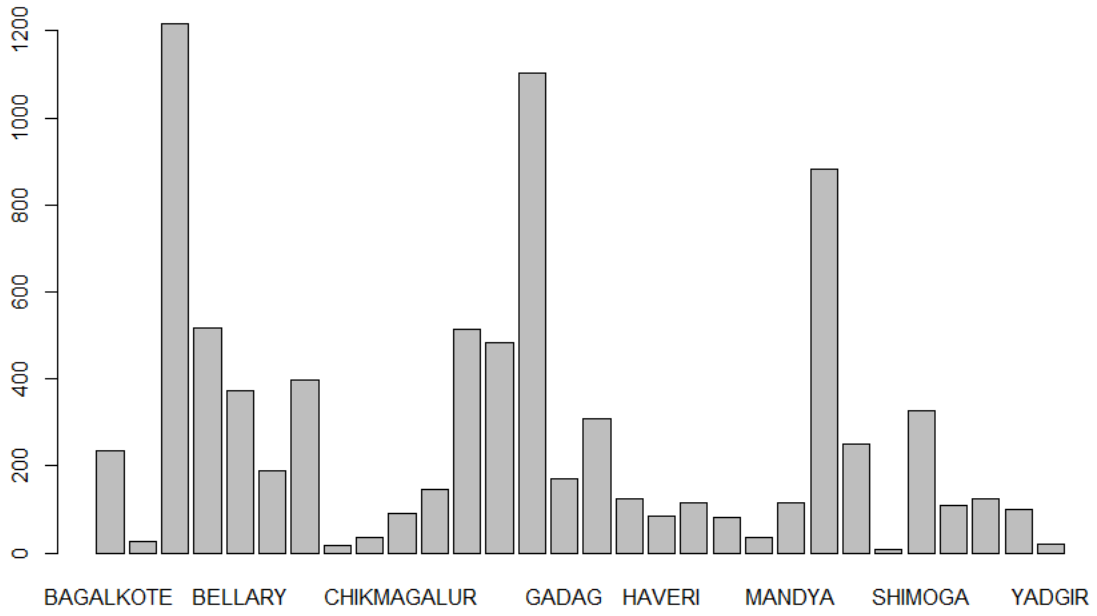
PieChart:

⇒ pie(b$Registered_Infant_Death)



BarGraph:

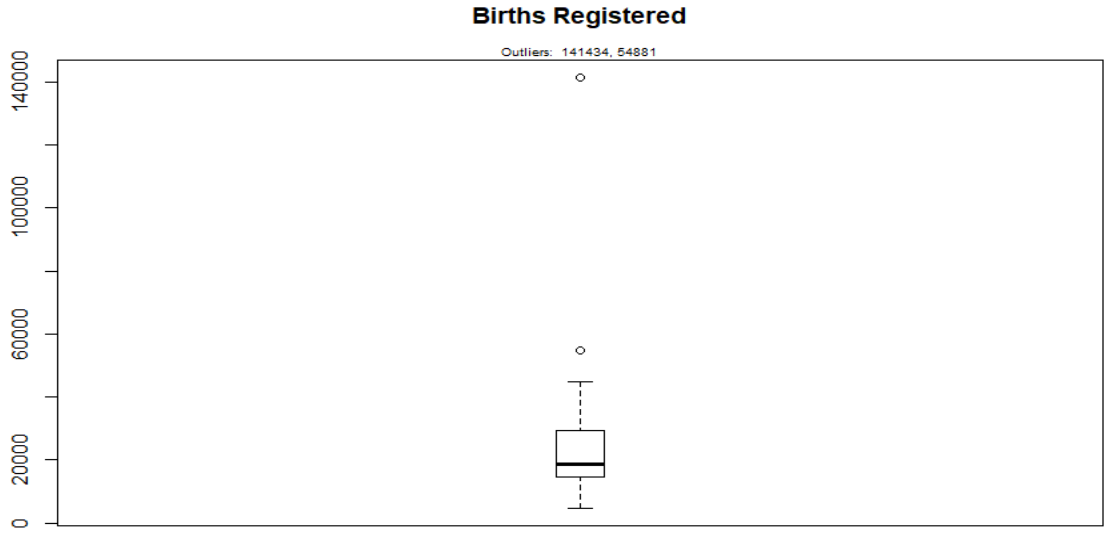
⇒ barplot(b$Registered_Infant_Death, names.arg = b$District)



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

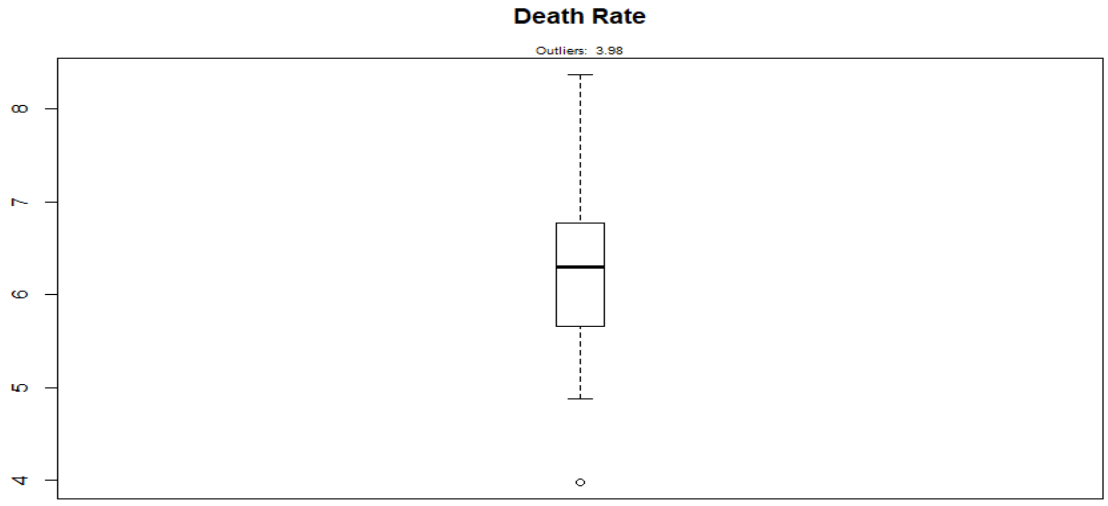
Outliers:

⇒ outlier1=boxplot.stats(b$Births_Registered)$out
⇒ boxplot(b$Births_Registered, main="Births Registered", boxwex=0.1)
⇒ mtext(paste("Outliers: ", paste(outlier1, collapse = ", ")),cex=0.6)
⇒

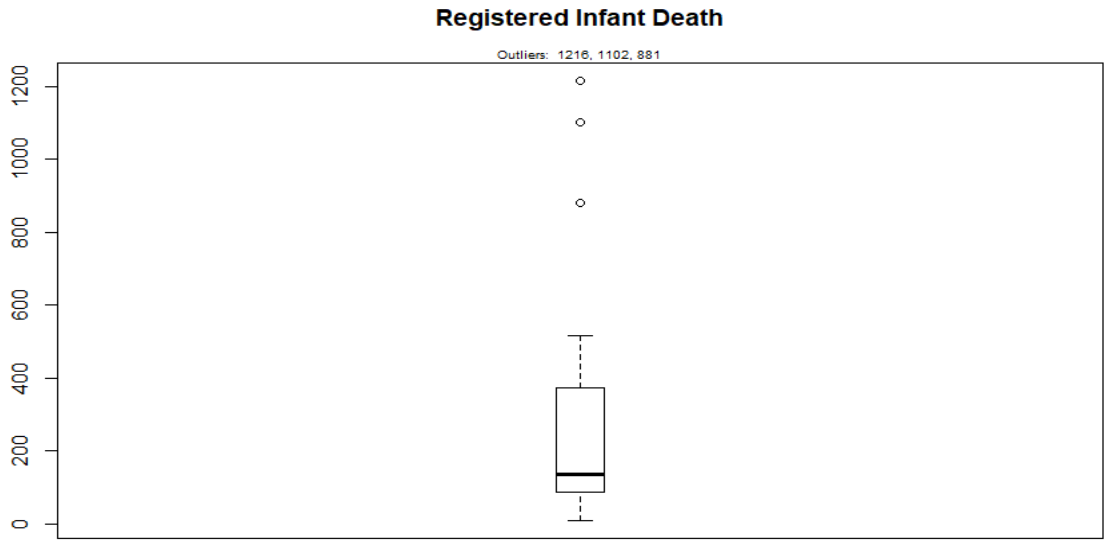


⇒ outlier2=boxplot.stats(b$Death_Registered)$out
⇒ boxplot(b$Death_Registered, main="Death Registered", boxwex=0.1)
⇒ mtext(paste("Outliers: ", paste(outlier2, collapse = ", ")),cex=0.6)

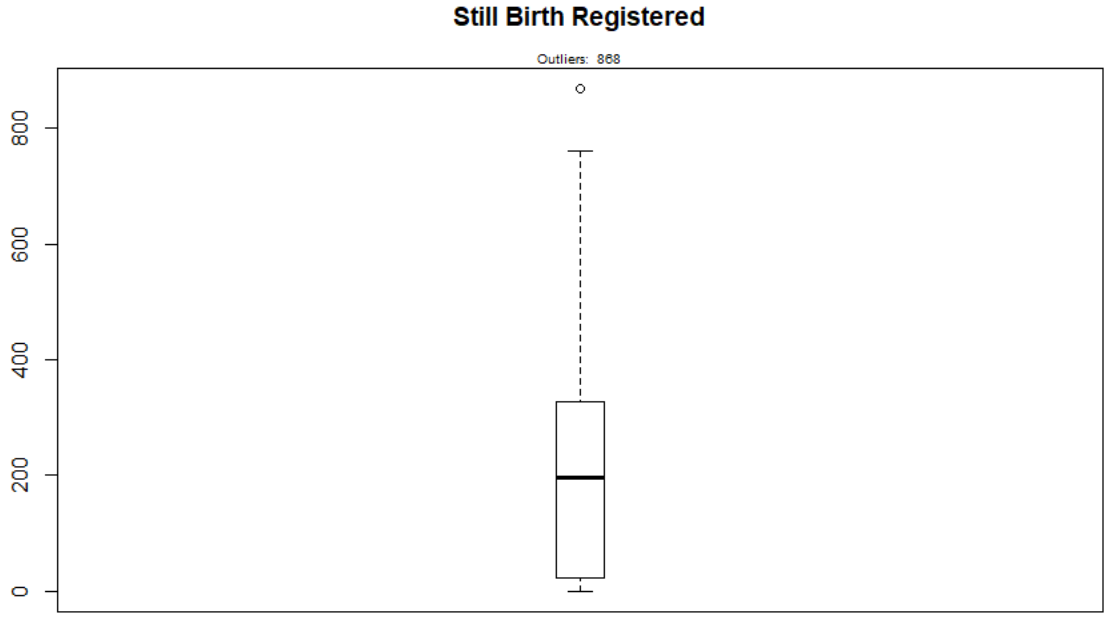
⇒ outlier3=boxplot.stats(b$Rate)$out
⇒ boxplot(b$Rate, main="Death Rate", boxwex=0.1)
⇒ mtext(paste("Outliers: ", paste(outlier3, collapse = ", ")),cex=0.6)



⇒ outlier4=boxplot.stats(b$Registered_Infant_Death)$out
⇒ boxplot(b$Registered_Infant_Death, main="Registered Infant Death", boxwex=0.1)
⇒ mtext(paste("Outliers: ", paste(outlier4, collapse = ", ")),cex=0.6)



⇒ outlier5=boxplot.stats(b$Still.Birth_Registered)$out
⇒ boxplot(b$Still.Birth_Registered, main="Still Birth Registered", boxwex=0.1)
⇒ mtext(paste("Outliers: ", paste(outlier5, collapse = ", ")),cex=0.6)



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

Plot of Births Registered in Urban and Rural Across 2011,2014,2015,2016:

⇒ library(ggplot2)

⇒ library(reshape2)

⇒ a=read.csv(file="G:\\Required\\6th Sem\\DA\\Lab\\Lab1\\all.csv", head = TRUE, sep = ",", nrows=30)

⇒ a

	District	X2011	X2014	X2015	X2016
1	BAGALKOTE	48759	48303	48851	47401
2	BANGALORE (R)	9292	11236	8965	10697
3	BANGALORE (U)	146105	154766	159272	160161
4	BELGAUM	96159	89606	89019	94861
5	BELLARY	54319	52110	54056	55388
6	BIDAR	40487	34170	36576	47339
7	BIJAPUR	61009	57550	52268	54769
8	CHAMARAJANAGAR	13733	11150	10395	11102
9	CHICKBALLAPUR	17918	14718	15039	16996
10	CHIKMAGALUR	17366	15077	15549	15005
11	CHITRADURGA	27130	24418	23741	24773
12	DAKSHINA KANNADA	38335	38264	39521	41383
13	DAVANAGERE	37149	40771	35424	34760
14	DHARWAD	39537	40351	38980	37364
15	GADAG	22567	20367	20007	19458
16	GULBARGA	49277	71232	58595	63707
17	HASSAN	26445	25999	25224	25452
18	HAVERI	30117	27154	24075	26453
19	KODAGU	8300	7808	7971	8179
20	KOLAR	21667	25401	23748	24614
21	KOPPAL	32610	26246	26978	28822
22	MANDYA	22280	20679	20107	19949
23	MYSORE	49105	45668	47374	54649
24	RAICHUR	31689	27624	25368	28902
25	RAMANAGAR	13547	11091	12036	10215
26	SHIMOGA	33555	31476	32992	33753
27	TUMKUR	41075	36189	36643	39942
28	UDUPI	19928	19558	18724	19356
29	UTTARA KANNADA	26425	26588	28806	25936
30	YADGIR	32677	31960	16944	25872

```
⇒ ggplot(sp, aes(x=District, y=value)) +
⇒   geom_bar(aes(fill=variable), stat="identity", position = "dodge") +
⇒   theme(axis.text.x = element_text(angle=75, hjust = 1))
```