

Veri Madenciliği Basit Komutlar

Gökhan Lülecı

28/03/2021

Giriş

Bu çalışma Veri Madenciliği basit komutları kapsamında hazırlanmıştır.

Ödev kapsamında yapılan uygulamalar için R içerisinde yer alan **state.x77** veri tabanı kullanılmıştır.

Öncelikle veriyi tanıyalım.

state.x77: Birleşik Devletlerin 50 Eyaletinin 8 Farklı Değişkene ait değerlerini göstermektedir.

Population: 1 Temmuz 1975 tarihi itibarıyla eyaletlerin mevcut nüfus toplamını göstermektedir.

Income: 1974 Yılı itibarıyla Kişi Başı Geliri göstermektedir. *Illiteracy:* 1970 yılındaki nüfusa göre okuma-yazma bilmeme oranını göstermektedir. *Life Exp:* 1969 ve 1971 yılları sürecinde ortalama ömür beklentisini göstermektedir. *Murder:* 1976 yılı itibarıyla 100.000 kişi başına düşen cinayet ve ihmal-dışı adam öldürme oranını göstermektedir. *HS Grad:* 1970 yılı itibarıyla Lise Mezunu oranını göstermektedir. *Frost:* Eyaletin başkentinde veya en büyük kentinde ölçülen minimum sıcaklığın donma noktasının altında olduğu ortalama gün sayısını göstermektedir. (1931–1960 arası) *Area:* Eyaletin yüzölçümünün mil-kare olarak karşılığıdır.

```
str(state.x77)
```

```
## 'data.frame':    50 obs. of  8 variables:
## $ Population: num  3615 365 2212 2110 21198 ...
## $ Income    : num  3624 6315 4530 3378 5114 ...
## $ Illiteracy: num   2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3 2 ...
## $ Life.Exp   : num   69 69.3 70.5 70.7 71.7 ...
## $ Murder     : num   15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2 10.7 13.9 ...
## $ HS.Grad    : num   41.3 66.7 58.1 39.9 62.6 63.9 56 54.6 52.6 40.6 ...
## $ Frost      : num    20 152 15 65 20 166 139 103 11 60 ...
## $ Area       : num  50708 566432 113417 51945 156361 ...
```

```
head(state.x77)
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20	50708
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
## Arkansas	2110	3378	1.9	70.66	10.1	39.9	65	51945
## California	21198	5114	1.1	71.71	10.3	62.6	20	156361
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766

Soru 1: Veri seti üzerinde her komut için 1 adet örnek işlemi için uygulayarak kısaca ne yaptığınızı belirtiniz.

1.1 [n,] komutu

Bu komut ile söz konusu eyaletler içinden n ‘inci satırda yer alan eyalete ait tüm veriler getirilir. Örneğin, **Massachusetts** Eyaleti için ilgili verileri getirelim.

labels(state.x77) ile veri incelendiğinde **Massachusetts** eyaletinin 21. Sırada olduğu gözükmemektedir. Söz konusu eyalete ilişkin bilgiler aşağıdaki kodlar vasıtası ile çağırılır.

```
state.x77[21,]
```

```
##              Population Income Illiteracy Life.Exp Murder HS.Grad Frost Area
## Massachusetts      5814   4755         1.1   71.83    3.3   58.5   103 7826
```

1.2 data[-n,] komutu

Bu komut ile söz konusu eyaletler içinden n ‘inci satırda yer alan tüm veriler çıkartılır. Örneğin, az önce incelediğimiz **Massachusetts** Eyaleti için ilgili verileri veritabanından kaldıralım ve satır sayısına bakalım.

```
nrow(state.x77)
```

```
## [1] 50
```

```
nrow(state.x77[-21,])
```

```
## [1] 49
```

1.3 data[1:n,] komutu

Bu komut ile söz konusu eyaletler içinden 1. sıradan itibaren istenen n ‘inci sıraya kadar yer alan tüm veriler seçilir. Örneğin, En baştan itibaren yer alan 10 eyaleti seçelim.

```
state.x77[1:10,]
```

```
##              Population Income Illiteracy Life.Exp Murder HS.Grad Frost Area
## Alabama          3615   3624         2.1   69.05   15.1   41.3    20 50708
## Alaska             365   6315         1.5   69.31   11.3   66.7   152 566432
## Arizona           2212   4530         1.8   70.55    7.8   58.1    15 113417
## Arkansas           2110   3378         1.9   70.66   10.1   39.9    65 51945
## California        21198  5114         1.1   71.71   10.3   62.6    20 156361
## Colorado           2541   4884         0.7   72.06    6.8   63.9   166 103766
## Connecticut        3100   5348         1.1   72.48    3.1   56.0   139  4862
## Delaware            579   4809         0.9   70.06    6.2   54.6   103  1982
## Florida            8277   4815         1.3   70.66   10.7   52.6    11 54090
## Georgia            4931   4091         2.0   68.54   13.9   40.6    60 58073
```

1.4 data[-(1:n),] komutu

Bu komut ile söz konusu eyaletler içinden 1. sıradan itibaren istenen n 'inci sıraya kadar yer alan tüm veriler çıkartılır. Örneğin, sadece ilk ve son 5 eyalet ile ilgilendiğimizde 6'dan başlayarak 44. eyalet dahil tüm eyaletleri veriden çıkartalım.

```
nrow(state.x77)
```

```
## [1] 50
```

```
state.x77[-(6:44),]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20	50708
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
## Arkansas	2110	3378	1.9	70.66	10.1	39.9	65	51945
## California	21198	5114	1.1	71.71	10.3	62.6	20	156361
## Vermont	472	3907	0.6	71.64	5.5	57.1	168	9267
## Virginia	4981	4701	1.4	70.08	9.5	47.8	85	39780
## Washington	3559	4864	0.6	71.72	4.3	63.5	32	66570
## West Virginia	1799	3617	1.4	69.48	6.7	41.6	100	24070
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149	54464
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203

1.5 data[c(i,j,k),] komutu

Bu komut ile belirlenen i, j ve k yerine satır numarası yazılarak seçilen 3 eyalete ait veriler çağırılır. Örneğin İlk(1) , Ortanca (25) ve sonuncu(50) eyaletleri ve bilgilerini çağıralım.

```
state.x77[c(1,25,50),]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20	50708
## Missouri	4767	4254	0.8	70.69	9.3	48.8	108	68995
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203

1.6 data[x > y,] komutu

Bu komut ile $x > y$ mantık testine uygun gelen satırlara ait tüm sütun verisi gelir. Örneğin yüzölçümü 100.000 inç kareden fazla olan eyaletlere ilişkin tüm bilgileri getirelim.

```
state.x77[state.x77$Area > 100000,]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
## California	21198	5114	1.1	71.71	10.3	62.6	20	156361
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766
## Montana	746	4347	0.6	70.56	5.0	59.2	155	145587
## Nevada	590	5149	0.5	69.03	11.5	65.2	188	109889
## New Mexico	1144	3601	2.2	70.32	9.7	55.2	120	121412
## Texas	12237	4188	2.2	70.90	12.2	47.4	35	262134

1.7 data[,m] komutu

Bu komut ile m'inci sütuna ilişkin tüm satır verisi getirilir. Örneğin tüm eyaletlere ait sadece kişi başı gelir rakamlarını getirelim.

```
colnames(state.x77)[2]
```

```
## [1] "Income"
```

```
state.x77[,2]
```

```
## [1] 3624 6315 4530 3378 5114 4884 5348 4809 4815 4091 4963 4119 5107 4458 4628
## [16] 4669 3712 3545 3694 5299 4755 4751 4675 3098 4254 4347 4508 5149 4281 5237
## [31] 3601 4903 3875 5087 4561 3983 4660 4449 4558 3635 4167 3821 4188 4022 3907
## [46] 4701 4864 3617 4468 4566
```

1.8 data[,-m] komutu

Bu komut ile m'inci sütununda yer alan tüm verisi çıkartılır. Örneğin eyaletlerde işlenen cinayetlere ilişkin verileri çıkaralım ve oluşan kolon sayısına bakalım.

```
ncol(state.x77)
```

```
## [1] 8
```

```
ncol(state.x77[,-5])
```

```
## [1] 7
```

1.9 data[1:m] komutu

Bu komut ile 1'den m'inci sütuna kadar yer alan tüm satırlara ilişkin bilgiler getirilir. Örneğin tüm eyaletlere ait Popülasyon(1), Kişi Başı Gelir(2) ve Okuma-Yazma Bilmeme Oranı(3) verilerini getirelim.

```
colnames(state.x77)[1:3]
```

```
## [1] "Population" "Income"      "Illiteracy"
```

```
state.x77[,1:3]
```

```
##           Population Income Illiteracy
## Alabama           3615   3624         2.1
## Alaska             365   6315         1.5
## Arizona           2212   4530         1.8
## Arkansas           2110   3378         1.9
## California        21198   5114         1.1
## Colorado           2541   4884         0.7
## Connecticut        3100   5348         1.1
## Delaware           579   4809         0.9
```

## Florida	8277	4815	1.3
## Georgia	4931	4091	2.0
## Hawaii	868	4963	1.9
## Idaho	813	4119	0.6
## Illinois	11197	5107	0.9
## Indiana	5313	4458	0.7
## Iowa	2861	4628	0.5
## Kansas	2280	4669	0.6
## Kentucky	3387	3712	1.6
## Louisiana	3806	3545	2.8
## Maine	1058	3694	0.7
## Maryland	4122	5299	0.9
## Massachusetts	5814	4755	1.1
## Michigan	9111	4751	0.9
## Minnesota	3921	4675	0.6
## Mississippi	2341	3098	2.4
## Missouri	4767	4254	0.8
## Montana	746	4347	0.6
## Nebraska	1544	4508	0.6
## Nevada	590	5149	0.5
## New Hampshire	812	4281	0.7
## New Jersey	7333	5237	1.1
## New Mexico	1144	3601	2.2
## New York	18076	4903	1.4
## North Carolina	5441	3875	1.8
## North Dakota	637	5087	0.8
## Ohio	10735	4561	0.8
## Oklahoma	2715	3983	1.1
## Oregon	2284	4660	0.6
## Pennsylvania	11860	4449	1.0
## Rhode Island	931	4558	1.3
## South Carolina	2816	3635	2.3
## South Dakota	681	4167	0.5
## Tennessee	4173	3821	1.7
## Texas	12237	4188	2.2
## Utah	1203	4022	0.6
## Vermont	472	3907	0.6
## Virginia	4981	4701	1.4
## Washington	3559	4864	0.6
## West Virginia	1799	3617	1.4
## Wisconsin	4589	4468	0.7
## Wyoming	376	4566	0.6

1.10 data[,-(1:m)] komutu

Bu komut ile 1'den m'inci sütuna kadar yer alan tüm kolonlar çıkartılır. Örneğin sadece eyaletlerde işlenen donma derecesinin altındaki gün sayısı (7) ve yüzölçümü (8) verilerini kontrol etmek için en baştan itibaren 6. sütuna kadarki tüm verileri çıkaralım.

```
colnames(state.x77)
```

```
## [1] "Population" "Income"      "Illiteracy" "Life.Exp"    "Murder"
## [6] "HS.Grad"    "Frost"       "Area"
```

```
state.x77[,-(1:6)]
```

##	Frost	Area
## Alabama	20	50708
## Alaska	152	566432
## Arizona	15	113417
## Arkansas	65	51945
## California	20	156361
## Colorado	166	103766
## Connecticut	139	4862
## Delaware	103	1982
## Florida	11	54090
## Georgia	60	58073
## Hawaii	0	6425
## Idaho	126	82677
## Illinois	127	55748
## Indiana	122	36097
## Iowa	140	55941
## Kansas	114	81787
## Kentucky	95	39650
## Louisiana	12	44930
## Maine	161	30920
## Maryland	101	9891
## Massachusetts	103	7826
## Michigan	125	56817
## Minnesota	160	79289
## Mississippi	50	47296
## Missouri	108	68995
## Montana	155	145587
## Nebraska	139	76483
## Nevada	188	109889
## New Hampshire	174	9027
## New Jersey	115	7521
## New Mexico	120	121412
## New York	82	47831
## North Carolina	80	48798
## North Dakota	186	69273
## Ohio	124	40975
## Oklahoma	82	68782
## Oregon	44	96184
## Pennsylvania	126	44966
## Rhode Island	127	1049
## South Carolina	65	30225
## South Dakota	172	75955
## Tennessee	70	41328
## Texas	35	262134
## Utah	137	82096
## Vermont	168	9267
## Virginia	85	39780
## Washington	32	66570
## West Virginia	100	24070
## Wisconsin	149	54464
## Wyoming	173	97203

1.11 data[,c(i,j,k)] komutu

Bu komut ile belirlenen i, j ve k yerine numara yazılarak seçilen 3 özelliğe ilişkin eyaletlerin tamamının verileri çağırılır. Örneğin Popülasyon(1) , Ortalama Yaşam Beklentisi (4) ve Lise Mezunu Oranı(6) değişkenlerine ait tüm eyaletler bilgilerini çağıralım.

```
colnames(state.x77)
```

```
## [1] "Population" "Income"      "Illiteracy" "Life.Exp"    "Murder"
## [6] "HS.Grad"     "Frost"        "Area"
```

```
state.x77[,c(1,4,6)]
```

##	Population	Life.Exp	HS.Grad
## Alabama	3615	69.05	41.3
## Alaska	365	69.31	66.7
## Arizona	2212	70.55	58.1
## Arkansas	2110	70.66	39.9
## California	21198	71.71	62.6
## Colorado	2541	72.06	63.9
## Connecticut	3100	72.48	56.0
## Delaware	579	70.06	54.6
## Florida	8277	70.66	52.6
## Georgia	4931	68.54	40.6
## Hawaii	868	73.60	61.9
## Idaho	813	71.87	59.5
## Illinois	11197	70.14	52.6
## Indiana	5313	70.88	52.9
## Iowa	2861	72.56	59.0
## Kansas	2280	72.58	59.9
## Kentucky	3387	70.10	38.5
## Louisiana	3806	68.76	42.2
## Maine	1058	70.39	54.7
## Maryland	4122	70.22	52.3
## Massachusetts	5814	71.83	58.5
## Michigan	9111	70.63	52.8
## Minnesota	3921	72.96	57.6
## Mississippi	2341	68.09	41.0
## Missouri	4767	70.69	48.8
## Montana	746	70.56	59.2
## Nebraska	1544	72.60	59.3
## Nevada	590	69.03	65.2
## New Hampshire	812	71.23	57.6
## New Jersey	7333	70.93	52.5
## New Mexico	1144	70.32	55.2
## New York	18076	70.55	52.7
## North Carolina	5441	69.21	38.5
## North Dakota	637	72.78	50.3
## Ohio	10735	70.82	53.2
## Oklahoma	2715	71.42	51.6
## Oregon	2284	72.13	60.0
## Pennsylvania	11860	70.43	50.2
## Rhode Island	931	71.90	46.4

## South Carolina	2816	67.96	37.8
## South Dakota	681	72.08	53.3
## Tennessee	4173	70.11	41.8
## Texas	12237	70.90	47.4
## Utah	1203	72.90	67.3
## Vermont	472	71.64	57.1
## Virginia	4981	70.08	47.8
## Washington	3559	71.72	63.5
## West Virginia	1799	69.48	41.6
## Wisconsin	4589	72.48	54.5
## Wyoming	376	70.29	62.9

1.12 data[,x > y] komutu

Bu komut ile $x > y$ mantık testine uygun gelen sütünlara ait tüm satır verileri getirilir. Örneğin 2.satırda 100.000 den az olan tüm kolonlardaki tüm satırlara ilişkin tüm bilgileri getirelim. Görüldüğü üzere Alaska'nın yüzölçümü 100.000'den büyük olduğu için söz konusu kolon çağırılmamıştır.

```
state.x77[,state.x77[2,] < 100000]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20
## Alaska	365	6315	1.5	69.31	11.3	66.7	152
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15
## Arkansas	2110	3378	1.9	70.66	10.1	39.9	65
## California	21198	5114	1.1	71.71	10.3	62.6	20
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166
## Connecticut	3100	5348	1.1	72.48	3.1	56.0	139
## Delaware	579	4809	0.9	70.06	6.2	54.6	103
## Florida	8277	4815	1.3	70.66	10.7	52.6	11
## Georgia	4931	4091	2.0	68.54	13.9	40.6	60
## Hawaii	868	4963	1.9	73.60	6.2	61.9	0
## Idaho	813	4119	0.6	71.87	5.3	59.5	126
## Illinois	11197	5107	0.9	70.14	10.3	52.6	127
## Indiana	5313	4458	0.7	70.88	7.1	52.9	122
## Iowa	2861	4628	0.5	72.56	2.3	59.0	140
## Kansas	2280	4669	0.6	72.58	4.5	59.9	114
## Kentucky	3387	3712	1.6	70.10	10.6	38.5	95
## Louisiana	3806	3545	2.8	68.76	13.2	42.2	12
## Maine	1058	3694	0.7	70.39	2.7	54.7	161
## Maryland	4122	5299	0.9	70.22	8.5	52.3	101
## Massachusetts	5814	4755	1.1	71.83	3.3	58.5	103
## Michigan	9111	4751	0.9	70.63	11.1	52.8	125
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160
## Mississippi	2341	3098	2.4	68.09	12.5	41.0	50
## Missouri	4767	4254	0.8	70.69	9.3	48.8	108
## Montana	746	4347	0.6	70.56	5.0	59.2	155
## Nebraska	1544	4508	0.6	72.60	2.9	59.3	139
## Nevada	590	5149	0.5	69.03	11.5	65.2	188
## New Hampshire	812	4281	0.7	71.23	3.3	57.6	174
## New Jersey	7333	5237	1.1	70.93	5.2	52.5	115
## New Mexico	1144	3601	2.2	70.32	9.7	55.2	120
## New York	18076	4903	1.4	70.55	10.9	52.7	82

## North Carolina	5441	3875	1.8	69.21	11.1	38.5	80
## North Dakota	637	5087	0.8	72.78	1.4	50.3	186
## Ohio	10735	4561	0.8	70.82	7.4	53.2	124
## Oklahoma	2715	3983	1.1	71.42	6.4	51.6	82
## Oregon	2284	4660	0.6	72.13	4.2	60.0	44
## Pennsylvania	11860	4449	1.0	70.43	6.1	50.2	126
## Rhode Island	931	4558	1.3	71.90	2.4	46.4	127
## South Carolina	2816	3635	2.3	67.96	11.6	37.8	65
## South Dakota	681	4167	0.5	72.08	1.7	53.3	172
## Tennessee	4173	3821	1.7	70.11	11.0	41.8	70
## Texas	12237	4188	2.2	70.90	12.2	47.4	35
## Utah	1203	4022	0.6	72.90	4.5	67.3	137
## Vermont	472	3907	0.6	71.64	5.5	57.1	168
## Virginia	4981	4701	1.4	70.08	9.5	47.8	85
## Washington	3559	4864	0.6	71.72	4.3	63.5	32
## West Virginia	1799	3617	1.4	69.48	6.7	41.6	100
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173

1.13 data[,c(1:m,i,j,k)] komutu

Bu komut ile belirlenen i, j ve k kolonlarının bir kopyaları 1:m ile belirtilen sütünlara kopya olarak eklenir. Örneğin 1:3 (İlk üç sütün)'a ek olarak Frost(6), Yüzölçümü(7) ve Lise Mezuniyeti(8) oranlarını ekleyelim. Böylece 4 ve 5. kolonları görüntüleyelim.

```
state.x77[,c(1:3,6,7,8)]
```

##	Population	Income	Illiteracy	HS.Grad	Frost	Area
## Alabama	3615	3624	2.1	41.3	20	50708
## Alaska	365	6315	1.5	66.7	152	566432
## Arizona	2212	4530	1.8	58.1	15	113417
## Arkansas	2110	3378	1.9	39.9	65	51945
## California	21198	5114	1.1	62.6	20	156361
## Colorado	2541	4884	0.7	63.9	166	103766
## Connecticut	3100	5348	1.1	56.0	139	4862
## Delaware	579	4809	0.9	54.6	103	1982
## Florida	8277	4815	1.3	52.6	11	54090
## Georgia	4931	4091	2.0	40.6	60	58073
## Hawaii	868	4963	1.9	61.9	0	6425
## Idaho	813	4119	0.6	59.5	126	82677
## Illinois	11197	5107	0.9	52.6	127	55748
## Indiana	5313	4458	0.7	52.9	122	36097
## Iowa	2861	4628	0.5	59.0	140	55941
## Kansas	2280	4669	0.6	59.9	114	81787
## Kentucky	3387	3712	1.6	38.5	95	39650
## Louisiana	3806	3545	2.8	42.2	12	44930
## Maine	1058	3694	0.7	54.7	161	30920
## Maryland	4122	5299	0.9	52.3	101	9891
## Massachusetts	5814	4755	1.1	58.5	103	7826
## Michigan	9111	4751	0.9	52.8	125	56817
## Minnesota	3921	4675	0.6	57.6	160	79289
## Mississippi	2341	3098	2.4	41.0	50	47296
## Missouri	4767	4254	0.8	48.8	108	68995

## Montana	746	4347	0.6	59.2	155	145587
## Nebraska	1544	4508	0.6	59.3	139	76483
## Nevada	590	5149	0.5	65.2	188	109889
## New Hampshire	812	4281	0.7	57.6	174	9027
## New Jersey	7333	5237	1.1	52.5	115	7521
## New Mexico	1144	3601	2.2	55.2	120	121412
## New York	18076	4903	1.4	52.7	82	47831
## North Carolina	5441	3875	1.8	38.5	80	48798
## North Dakota	637	5087	0.8	50.3	186	69273
## Ohio	10735	4561	0.8	53.2	124	40975
## Oklahoma	2715	3983	1.1	51.6	82	68782
## Oregon	2284	4660	0.6	60.0	44	96184
## Pennsylvania	11860	4449	1.0	50.2	126	44966
## Rhode Island	931	4558	1.3	46.4	127	1049
## South Carolina	2816	3635	2.3	37.8	65	30225
## South Dakota	681	4167	0.5	53.3	172	75955
## Tennessee	4173	3821	1.7	41.8	70	41328
## Texas	12237	4188	2.2	47.4	35	262134
## Utah	1203	4022	0.6	67.3	137	82096
## Vermont	472	3907	0.6	57.1	168	9267
## Virginia	4981	4701	1.4	47.8	85	39780
## Washington	3559	4864	0.6	63.5	32	66570
## West Virginia	1799	3617	1.4	41.6	100	24070
## Wisconsin	4589	4468	0.7	54.5	149	54464
## Wyoming	376	4566	0.6	62.9	173	97203

1.14 data[x > y,a != b] komutu

Bu komut ile (x > y) koşulunu sağlayan satırlar ile (a != b) koşulunu sağlayan sütunlar seçilerek getirilir. Önce getirilen seçeneklere bakalım sonra formülü kullanarak sadece istenen verileri getirelim. Örneğin Nüfus'u 200.000 'in üzerinde olan eyaletlerin donma sıcaklığın altındaki gün sayılarını getirelim.

```
state.x77[state.x77$Area > 200000 , ]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Texas	12237	4188	2.2	70.90	12.2	47.4	35	262134

```
state.x77[state.x77$Area > 200000 , "Frost" ]
```

```
## [1] 152 35
```

1.15 data[c(1:n,i,j,k),] komutu

Bu komut ile i, j, ve k satırlarının bir kopyası 1'den n'e kadar seçilen satırların üzerine eklenir. Örnek olarak 1'den 3'ye kadar olan (İlk 3) eyaletin üzerine son 3 (48,49,50) eyaletlerini ekleyelim.

```
state.x77[c(1:3,48,49,50),]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20	50708

## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
## West Virginia	1799	3617	1.4	69.48	6.7	41.6	100	24070
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149	54464
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203

Soru 2: Verilen veri setinde kendi seçeceğimiz 3 değişkeninin medyandan büyük değerlerini sıralı olarak gösteren R komutunu oluşturarak sonucu gösterelim.

Aynı işlemleri bir de medyan yerine aritmetik ortalama kullanarak tekrarlayalım. Medyan ve Aritmetik ortalama göre farklılıkları irdeleyelim.

Söz konusu inceleme için Nüfus(Population), Gelir(Income) ve Yüzölçümü(Area) değişkenlerini inceleyelim.

3 Değişken için Medyan ve Ortalama değerlerine göre filtreleme yapıp sıralama

Bu komut ile söz konusu seçilen değişkenleri istenen koşullara göre sıralayacağız. Öncelikle popülasyona ait ortalama (mean) ve medyan (median) değerlerini buluyoruz. Ardından söz konusu medyan ve ortalama değerlerinden büyük olan eyaletleri ayrı ayrı listeleyip ardından büyükten küçüğe azaltacak sıralıyoruz.

Ortalama, bir verinin toplam değerlerinin toplam sayısının değerlerin sayısına bölünmesiyle bulunur iken; Medyan, bir veri küçükten büyüğe sıralandığında ortada yer alan değeridir. Verilerin incelenmesi için ortalama ya da ortanca(median) seçimi, verilerin türüne ve sonuç gereksinimine bağlıdır, çünkü aşırılıkların(outliers) görüldüğü bazı durumlarda median değeri ortalama daha iyi sonuçlar verir. Söz konusu örnekle aritmetik ortalama medyandan daha büyük bir değere sahip olduğu için , mean'e göre hesaplanan listede daha az sayıda eyalet yer almaktadır. Aykırı verileri olup olmadığı verinin normallik testlerine tabii tutulması ile gerçekleştirilebilir.

```
mean(state.x77$Population)
```

```
## [1] 4246.42
```

```
median(state.x77$Population)
```

```
## [1] 2838.5
```

```
pop_ortalama <- as.data.frame(state.x77[state.x77$Population > mean(state.x77$Population),])
pop_medyan <- as.data.frame(state.x77[state.x77$Population > median(state.x77$Population),])
```

```
nrow(pop_ortalama)
```

```
## [1] 16
```

```
nrow(pop_medyan)
```

```
## [1] 25
```

```
pop_medyan[with(pop_medyan, order(Population, decreasing = T)), ]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost
## California	21198	5114	1.1	71.71	10.3	62.6	20
## New York	18076	4903	1.4	70.55	10.9	52.7	82
## Texas	12237	4188	2.2	70.90	12.2	47.4	35
## Pennsylvania	11860	4449	1.0	70.43	6.1	50.2	126
## Illinois	11197	5107	0.9	70.14	10.3	52.6	127
## Ohio	10735	4561	0.8	70.82	7.4	53.2	124
## Michigan	9111	4751	0.9	70.63	11.1	52.8	125
## Florida	8277	4815	1.3	70.66	10.7	52.6	11
## New Jersey	7333	5237	1.1	70.93	5.2	52.5	115
## Massachusetts	5814	4755	1.1	71.83	3.3	58.5	103
## North Carolina	5441	3875	1.8	69.21	11.1	38.5	80
## Indiana	5313	4458	0.7	70.88	7.1	52.9	122
## Virginia	4981	4701	1.4	70.08	9.5	47.8	85
## Georgia	4931	4091	2.0	68.54	13.9	40.6	60
## Missouri	4767	4254	0.8	70.69	9.3	48.8	108
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149
## Tennessee	4173	3821	1.7	70.11	11.0	41.8	70
## Maryland	4122	5299	0.9	70.22	8.5	52.3	101
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160
## Louisiana	3806	3545	2.8	68.76	13.2	42.2	12
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20
## Washington	3559	4864	0.6	71.72	4.3	63.5	32
## Kentucky	3387	3712	1.6	70.10	10.6	38.5	95
## Connecticut	3100	5348	1.1	72.48	3.1	56.0	139
## Iowa	2861	4628	0.5	72.56	2.3	59.0	140
##	Area						
## California	156361						
## New York	47831						
## Texas	262134						
## Pennsylvania	44966						
## Illinois	55748						
## Ohio	40975						
## Michigan	56817						
## Florida	54090						
## New Jersey	7521						
## Massachusetts	7826						
## North Carolina	48798						
## Indiana	36097						
## Virginia	39780						
## Georgia	58073						
## Missouri	68995						
## Wisconsin	54464						
## Tennessee	41328						
## Maryland	9891						
## Minnesota	79289						
## Louisiana	44930						
## Alabama	50708						
## Washington	66570						
## Kentucky	39650						
## Connecticut	4862						
## Iowa	55941						

```
pop_ortalama[with(pop_ortalama, order(Population, decreasing = T)), ]
```

```
##           Population Income Illiteracy Life.Exp Murder HS.Grad Frost
## California      21198   5114         1.1   71.71   10.3   62.6    20
## New York        18076   4903         1.4   70.55   10.9   52.7    82
## Texas           12237   4188         2.2   70.90   12.2   47.4    35
## Pennsylvania    11860   4449         1.0   70.43    6.1   50.2   126
## Illinois        11197   5107         0.9   70.14   10.3   52.6   127
## Ohio            10735   4561         0.8   70.82    7.4   53.2   124
## Michigan         9111   4751         0.9   70.63   11.1   52.8   125
## Florida          8277   4815         1.3   70.66   10.7   52.6    11
## New Jersey       7333   5237         1.1   70.93    5.2   52.5   115
## Massachusetts   5814   4755         1.1   71.83    3.3   58.5   103
## North Carolina   5441   3875         1.8   69.21   11.1   38.5    80
## Indiana          5313   4458         0.7   70.88    7.1   52.9   122
## Virginia         4981   4701         1.4   70.08    9.5   47.8    85
## Georgia          4931   4091         2.0   68.54   13.9   40.6    60
## Missouri         4767   4254         0.8   70.69    9.3   48.8   108
## Wisconsin       4589   4468         0.7   72.48    3.0   54.5   149
##           Area
## California    156361
## New York      47831
## Texas         262134
## Pennsylvania  44966
## Illinois      55748
## Ohio          40975
## Michigan      56817
## Florida       54090
## New Jersey    7521
## Massachusetts 7826
## North Carolina 48798
## Indiana       36097
## Virginia      39780
## Georgia       58073
## Missouri      68995
## Wisconsin    54464
```

Gelire ait ortalama (mean) ve medyan (median) değerlerini buluyoruz. Ardından söz konusu medyan ve ortalama değerlerinden büyük olan eyaletleri ayrı ayrı listeleyp ardından büyükten küçüğe azaltacak sıralıyoruz.

Ortalama, bir verinin toplam değerlerinin toplam sayısının değerlerin sayısına bölünmesiyle bulunur iken; Medyan, bir veri küçükten büyüğe sıralandığında ortada yer alan değeridir. Verilerin incelenmesi için ortalama ya da ortanca(median) seçimi, verilerin türüne ve sonuç gereksinimine bağlıdır, çünkü aşırılıkların görüldüğü(outliers) bazı durumlarda ortalama ortalamanın üzerinde veya daha iyisi sonuç verir. Söz konusu örnekle aritmetik ortalama medyandan daha küçük bir değere sahip olduğu için , mean'e göre hesaplanan listede daha çok sayıda eyalet yer almaktadır. Aykırı verileri olup olmadığı verinin normallik testlerine tabii tutulması ile gerçekleştirilebilir.

```
mean(state.x77$Income)
```

```
## [1] 4435.8
```

```
median(state.x77$Income)
```

```
## [1] 4519
```

```
inc_ortalama <- as.data.frame(state.x77[state.x77$Income > mean(state.x77$Income),])
inc_medyan <- as.data.frame(state.x77[state.x77$Income > median(state.x77$Income),])

nrow(inc_ortalama)
```

```
## [1] 29
```

```
nrow(inc_medyan)
```

```
## [1] 25
```

```
inc_medyan[with(inc_medyan, order(Income, decreasing = T)), ]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Connecticut	3100	5348	1.1	72.48	3.1	56.0	139	4862
## Maryland	4122	5299	0.9	70.22	8.5	52.3	101	9891
## New Jersey	7333	5237	1.1	70.93	5.2	52.5	115	7521
## Nevada	590	5149	0.5	69.03	11.5	65.2	188	109889
## California	21198	5114	1.1	71.71	10.3	62.6	20	156361
## Illinois	11197	5107	0.9	70.14	10.3	52.6	127	55748
## North Dakota	637	5087	0.8	72.78	1.4	50.3	186	69273
## Hawaii	868	4963	1.9	73.60	6.2	61.9	0	6425
## New York	18076	4903	1.4	70.55	10.9	52.7	82	47831
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766
## Washington	3559	4864	0.6	71.72	4.3	63.5	32	66570
## Florida	8277	4815	1.3	70.66	10.7	52.6	11	54090
## Delaware	579	4809	0.9	70.06	6.2	54.6	103	1982
## Massachusetts	5814	4755	1.1	71.83	3.3	58.5	103	7826
## Michigan	9111	4751	0.9	70.63	11.1	52.8	125	56817
## Virginia	4981	4701	1.4	70.08	9.5	47.8	85	39780
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160	79289
## Kansas	2280	4669	0.6	72.58	4.5	59.9	114	81787
## Oregon	2284	4660	0.6	72.13	4.2	60.0	44	96184
## Iowa	2861	4628	0.5	72.56	2.3	59.0	140	55941
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203
## Ohio	10735	4561	0.8	70.82	7.4	53.2	124	40975
## Rhode Island	931	4558	1.3	71.90	2.4	46.4	127	1049
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417

```
inc_ortalama[with(inc_ortalama, order(Income, decreasing = T)), ]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Connecticut	3100	5348	1.1	72.48	3.1	56.0	139	4862
## Maryland	4122	5299	0.9	70.22	8.5	52.3	101	9891

## New Jersey	7333	5237	1.1	70.93	5.2	52.5	115	7521
## Nevada	590	5149	0.5	69.03	11.5	65.2	188	109889
## California	21198	5114	1.1	71.71	10.3	62.6	20	156361
## Illinois	11197	5107	0.9	70.14	10.3	52.6	127	55748
## North Dakota	637	5087	0.8	72.78	1.4	50.3	186	69273
## Hawaii	868	4963	1.9	73.60	6.2	61.9	0	6425
## New York	18076	4903	1.4	70.55	10.9	52.7	82	47831
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766
## Washington	3559	4864	0.6	71.72	4.3	63.5	32	66570
## Florida	8277	4815	1.3	70.66	10.7	52.6	11	54090
## Delaware	579	4809	0.9	70.06	6.2	54.6	103	1982
## Massachusetts	5814	4755	1.1	71.83	3.3	58.5	103	7826
## Michigan	9111	4751	0.9	70.63	11.1	52.8	125	56817
## Virginia	4981	4701	1.4	70.08	9.5	47.8	85	39780
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160	79289
## Kansas	2280	4669	0.6	72.58	4.5	59.9	114	81787
## Oregon	2284	4660	0.6	72.13	4.2	60.0	44	96184
## Iowa	2861	4628	0.5	72.56	2.3	59.0	140	55941
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203
## Ohio	10735	4561	0.8	70.82	7.4	53.2	124	40975
## Rhode Island	931	4558	1.3	71.90	2.4	46.4	127	1049
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
## Nebraska	1544	4508	0.6	72.60	2.9	59.3	139	76483
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149	54464
## Indiana	5313	4458	0.7	70.88	7.1	52.9	122	36097
## Pennsylvania	11860	4449	1.0	70.43	6.1	50.2	126	44966

Yüzölçümüne ait ortalama (mean) ve medyan (median) değerlerini buluyoruz. Ardından söz konusu medyan ve ortalama değerlerinden büyük olan eyaletleri ayrı ayrı listeleyip ardından büyükten küçüğe azaltacak sıralıyoruz.

Ortalama, bir verinin toplam değerlerinin toplam sayısının değerlerin sayısına bölünmesiyle bulunur iken; Medyan, bir veri küçükten büyüğe sıralandığında ortada yer alan değeridir. Verilerin incelenmesi için ortalama ya da ortanca(median) seçimi, verilerin türüne ve sonuç gereksinimine bağlıdır, çünkü aşırılıkların görüldüğü(outliers) bazı durumlarda ortalama ortalamamın üzerinde veya daha iyisi sonuç verir. Söz konusu örneklerle aritmetik ortalama medyandan daha küçük bir değere sahip olduğu için , mean'e göre hesaplanan listede daha çok sayıda eyalet yer almaktadır. Aykırı verileri olup olmadığı verinin normallik testlerine tabii tutulması ile gerçekleştirilebilir.

```
mean(state.x77$Area)
```

```
## [1] 70735.88
```

```
median(state.x77$Area)
```

```
## [1] 54277
```

```
area_ortalama <- as.data.frame(state.x77[state.x77$Area > mean(state.x77$Area),])
area_medyan <- as.data.frame(state.x77[state.x77$Area > median(state.x77$Area),])
```

```
nrow(area_ortalama)
```

```
## [1] 16
```

```
nrow(area_medyan)
```

```
## [1] 25
```

```
area_medyan[with(area_medyan, order(Area, decreasing = T)), ]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Texas	12237	4188	2.2	70.90	12.2	47.4	35	262134
## California	21198	5114	1.1	71.71	10.3	62.6	20	156361
## Montana	746	4347	0.6	70.56	5.0	59.2	155	145587
## New Mexico	1144	3601	2.2	70.32	9.7	55.2	120	121412
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
## Nevada	590	5149	0.5	69.03	11.5	65.2	188	109889
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203
## Oregon	2284	4660	0.6	72.13	4.2	60.0	44	96184
## Idaho	813	4119	0.6	71.87	5.3	59.5	126	82677
## Utah	1203	4022	0.6	72.90	4.5	67.3	137	82096
## Kansas	2280	4669	0.6	72.58	4.5	59.9	114	81787
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160	79289
## Nebraska	1544	4508	0.6	72.60	2.9	59.3	139	76483
## South Dakota	681	4167	0.5	72.08	1.7	53.3	172	75955
## North Dakota	637	5087	0.8	72.78	1.4	50.3	186	69273
## Missouri	4767	4254	0.8	70.69	9.3	48.8	108	68995
## Oklahoma	2715	3983	1.1	71.42	6.4	51.6	82	68782
## Washington	3559	4864	0.6	71.72	4.3	63.5	32	66570
## Georgia	4931	4091	2.0	68.54	13.9	40.6	60	58073
## Michigan	9111	4751	0.9	70.63	11.1	52.8	125	56817
## Iowa	2861	4628	0.5	72.56	2.3	59.0	140	55941
## Illinois	11197	5107	0.9	70.14	10.3	52.6	127	55748
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149	54464

```
area_ortalama[with(area_ortalama, order(Area, decreasing = T)), ]
```

##	Population	Income	Illiteracy	Life.Exp	Murder	HS.Grad	Frost	Area
## Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
## Texas	12237	4188	2.2	70.90	12.2	47.4	35	262134
## California	21198	5114	1.1	71.71	10.3	62.6	20	156361
## Montana	746	4347	0.6	70.56	5.0	59.2	155	145587
## New Mexico	1144	3601	2.2	70.32	9.7	55.2	120	121412
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
## Nevada	590	5149	0.5	69.03	11.5	65.2	188	109889
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203
## Oregon	2284	4660	0.6	72.13	4.2	60.0	44	96184
## Idaho	813	4119	0.6	71.87	5.3	59.5	126	82677
## Utah	1203	4022	0.6	72.90	4.5	67.3	137	82096
## Kansas	2280	4669	0.6	72.58	4.5	59.9	114	81787
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160	79289
## Nebraska	1544	4508	0.6	72.60	2.9	59.3	139	76483
## South Dakota	681	4167	0.5	72.08	1.7	53.3	172	75955

Soru 3: Veri setinde yer alan 2 değişken için özetleyici istatistikleri gösteren R komutunu oluşturarak sonuçları yorumlayalım.

Aşağıdaki kodlar ile verilere ilişkin Tanımlayıcı İstatistik Değerleri kontrol edilmiş ve özetleri sunulmuştur. Her iki tanımlayıcı istatistik verilerinde yer alan Mean , Median gibi temel göstergelerin yanı sıra Skewness ve Kurtosis gibi normallik göstergeleride yorumlanmıştır. Skewness değerler verinin Eğiklik/Çarpıklık düzeyini ifade ederken, Kurtosis ise Basıklık düzeyini anlatmaktadır. İlgili değerlerin istatistiklerinin yanı sıra standart hataları yer almaktadır bu standart hatalar, örneklemelerin ortalamasının, popülasyonun ortalamasından sapma derecesini ifade etmektedir.

Grafikler ve istatistikler ile verileri yorumlamadan önce söz konusu görselleştirme yapmak amacıyla “**psych**” ve “**ggplot2**” paketlerini kütüphaneden çağıralım.

```
library("psych")
library("ggplot2")
describe(state.x77$Income)
```

```
##      vars  n   mean      sd median trimmed   mad  min  max range skew kurtosis
## X1      1 50 4435.8 614.47   4519 4430.08 581.18 3098 6315  3217  0.2    0.24
##      se
## X1 86.9
```

```
summary(state.x77$Income)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      3098    3993    4519    4436    4814    6315
```

```
par(mfrow=c(1,2))
boxplot(state.x77$Income , col="990066", xlab="Income" )
hist(state.x77$Income, main = "Gelir Histogramı", xlab = "Dollar",col="990066")
```



Seçilmiş olan Gelir Dağılımına ilişkin betimleyici istatistikler ve grafikler sunulmuştur. Gelir değişkeni 50 adet numeric data'dan oluşmaktadır ve eyaletlerin ortalama dağılımlarını göstermektedir. Tüm eyaletlerin aritmetik ortalama geliri 4,435.8 Dollar iken Median'ı 4519 Dolardır. Standart Sapması 614.47 Dolar olan Gelir değişkeninin normallik kontrollerini yapmadan önce çeyrek değerlerine ve max min değerlerine bakalım. Verideki en düşük gelir 3098 dolar iken, en yüksek gelir 6315 Dolardır. Kutu-Bıyık grafiğinde görülebileceği üzere bu değer Outlier olarak tanımlanmıştır. En büyük ve en küçük veri arasındaki fark 3217 iken, Bu değerlere ilişkin Standart Hata 86.9'dur. Son olarak Ortanca Mutlak Sapma beklenildiği üzere standart sapmadan küçük ve 581.18'dir.

Normallik detay testleri yapılmadan histogram ve kutu bıyık grafiği çizilmiş ve veriye bakış atılmıştır. Buradan çıkarılacak sonuç verinin Quartile'ları arasında çok sayıda veri olduğu ve belki sağa çarpık olduğudur. Normallik analizleri takip eden grafiklerde detaylı olarak hesaplanacak bu tespitler irdelenecektir.

- Skewness: Popülasyon ile çalışıldığı için (ABD'deki tüm eyaletler) dağılımın çarpıklığı pozitif/negatif değerine göre sağa/sola çarpık olduğu sonuca, 1 ve 0 ile olan yakınlık uzaklık ilişkine göre ise çarpık olup olmadığı (çarpıklığın şiddetine) kanısına varılması gerekmektedir. Ayrıca R'da sonuçları 3 düzeltme katsayını düzelterek vermesi nedeniyle, Değerlerin 0'dan yakınlık ve uzaklıklarına bakmamız gerekmektedir. İlgili gelir değişkeni için SKewness değeri + 0.2 çıkmıştır. Dolayısı ile Histogram şeklinde de teyit edilebileceği üzere dağılım normale simetrik olarak kabul edilebilir fakat detay vermek gerekirse sağa orta derecede çarpık olduğu söylenebilir. Parametrik testleri ile analiz edilebilir durumdadır.

- Kurtosis: Normal dağılımın basıklık değeri 3 olduğu gerçeği ile hareket edilerek, fakat R 'ın sonuçları 3 düzeltme katsayını düzelterek vermesi nedeniyle, Değerlerin 0'dan yakınlık ve uzaklıklarına bakmamız gerekmektedir. Hesaplanan Kurtosis değeri 0.24 olarak hesaplandığı için, bu dağılımın Normal Dağılıma simetrik olduğu yorumu yapılabilir.

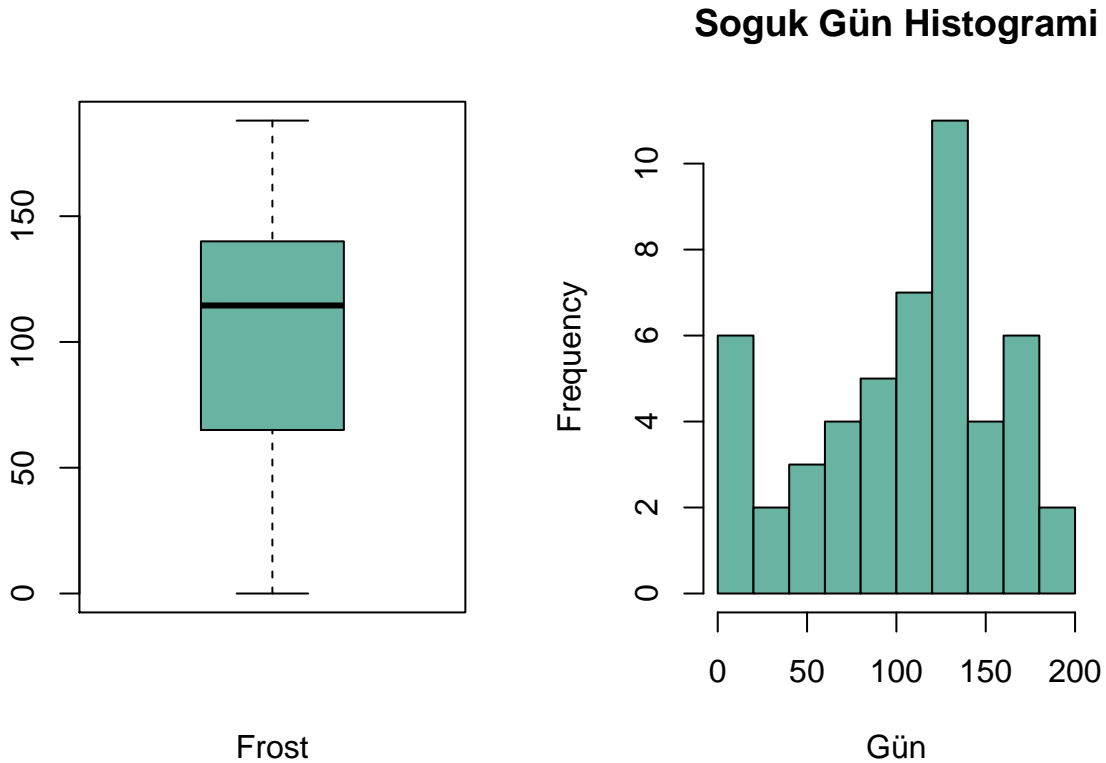
```
library("psych")
library("ggplot2")
describe(state.x77$Frost)
```

```
##      vars  n   mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1      1 50 104.46 51.98  114.5   106.8 53.37   0 188   188 -0.37   -0.94 7.35
```

```
summary(state.x77$Frost)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   66.25   114.50   104.46   139.75   188.00
```

```
par(mfrow=c(1,2))
boxplot(state.x77$Frost , col="#69b3a2", xlab="Frost" )
hist(state.x77$Frost, main = "Soğuk Gün Histogramı", xlab = "Gün",col="#69b3a2")
```



Seçilmiş olan Sıfırın altındaki Günler (Soğuk Gün olarak anılacaktır) Dağılımına ilişkin betimleyici istatistikler ve grafikler sunulmuştur. Soğuk gün değişkeni 50 adet numeric data'dan oluşmaktadır ve eyaletlerin ortalama dağılımlarını göstermektedir. Tüm eyaletlerin aritmetik ortalama olarak soğuk gün sayısı 104.46 Gün iken Median'ı 114.5 gündür. Standart Sapması 51.98 Gün olan Soğuk gün değişkeninin normallik kontrollerini yapmadan önce çeyrek değerlerine ve max min değerlerine bakalım. Verideki en az soğuk güne sahip eyalet 0 güne sahip iken, en yüksek soğuk güne sahip eyalet 188 güne sahiptir. En büyük ve en küçük veri arasındaki fark dolayısı ile 188 iken Bu değerlere ilişkin Standart Hata 7.35'dir. Son olarak Ortanca Mutlak Sapma (median - mena değerlerine bakılarak) beklenildiği üzere standart sapmadan büyük ve 53.37'dir. Kutu-Bıyık grafiğinde görülebileceği veride Outlier değer bulunmamaktadır.

Normallik detay testleri yapılmadan histogram ve kutu bıyık grafiğı çizilmiş ve veriye bakış atılmıştır. Buradan çıkarılacak sonuç verinin Quartile'ları arasında çok sayıda veri olduğu ve belki sola çarpık olduğudur. Normallik analizleri takip eden grafiklerde detaylı olarak hesaplanacak bu tespitler irdelenecektir.

- Skewness: Popölasyon ile çalışıldığı için (ABD'deki tüm eyaletler) dağılımın çarpıklığı pozitif/negatif değerine göre sağ/sola çarpık olduğu sonuca, 1 ve 0 ile olan yakınlık uzaklık ilişkine göre ise çarpıklık şiddetine varılması gerekmektedir. Ayrıca R'da sonuçları 3 düzeltme katsayını düzelterek vermesi nedeniyle, Değerlerin 0'dan yakınlık ve uzaklıklarına bakmamız gerekmektedir. İlgili gelir değişkeni için SKewness değeri -0.37 çıkmıştır. Dolayısı ile Histogram şeklinde de teyit edilebileceğı üzere dağılım normale simetrik olarak kabul edilebilir fakat detay vermek gerekirse sola orta derecede çarpık olduğu söylenebilir. Parametrik testleri ile analiz edilebilir durumdadır.

- Kurtosis: Normal dağılımın basıklık değeri 3 olduğu gerçeğı ile hareket edilerek, fakat R 'ın sonuçları 3 düzeltme katsayını düzelterek vermesi nedeniyle, Değerlerin 0'dan yakınlık ve uzaklıklarına bakmamız gerekmektedir. Hesaplanan Kurtosis değeri 0.24 olarak hesaplandığı için, bu dağılımın Normal Dağılıma simetrik olduğu yorumu yapılabilir fakat detay verilmesi istenirse çok az Platykürtik yada Basık olduğu söylenebilir. Yukarıdaki Histogram grafiğinde yer aldığı üzere, normal dağılıma oranla, dağılımın kuyrukları daha kısa ve daha incedir. Ayrıca merkez piki diğer verilere oranla daha düşük ve daha geniştir. Parametrik testleri ile analiz edilebilir durumdadır.