Seminor - week 3

Noriko Magara

25 September, 2015

Violent Crime Rates by US State

Preset dataset

I used "**USArrests**" dataset that contains the number of crimes per 100,000 people in each state in **1973**. The three types of crimes are: **Murder**, **Assault**, and **Rape**, as well as **Urban population**.

Below is the summary statistics of the dataset. The first command `data("USArrests")` loads the preset dataset. The second command `summary(USArrests)` gives summary statistics as shown below.

```
data("USArrests")
summary(USArrests)
```

```
##      Murder          Assault         UrbanPop          Rape
##  Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30
##  1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07
##  Median : 7.250   Median :159.0   Median :66.00   Median :20.10
##  Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23
##  3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18
##  Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00
```

Combining with more vectors

I also used "**state**" dataset that contains informations related to the 50 states. I combined "**state.abb**" and "**state.x77**" to the **USArrests** dataframe and created new one called "**UScombined**" that now has abbrebiations, population, income, etc.

```
# abbrebiations
ABB <- c(state.abb)
# demographics (population, income, ect.)
STATESFACT <- data.frame(state.x77)

# new dataframe w/ demographics
UScombined <- cbind(ABB, USArrests, STATESFACT)
```

The table below shows the whole dataframe.

	ABB	Murder	Assault	UrbanPop	Rape	Population	Income	Illiteracy	Life.Exp	Murder
Alabama	AL	13.2	236	58	21.2	3615	3624	2.1	69.05	15.1
Alaska	AK	10.0	263	48	44.5	365	6315	1.5	69.31	11.3
Arizona	AZ	8.1	294	80	31.0	2212	4530	1.8	70.55	7.8
Arkansas	AR	8.8	190	50	19.5	2110	3378	1.9	70.66	10.1
California	CA	9.0	276	91	40.6	21198	5114	1.1	71.71	10.3
Colorado	CO	7.9	204	78	38.7	2541	4884	0.7	72.06	6.8

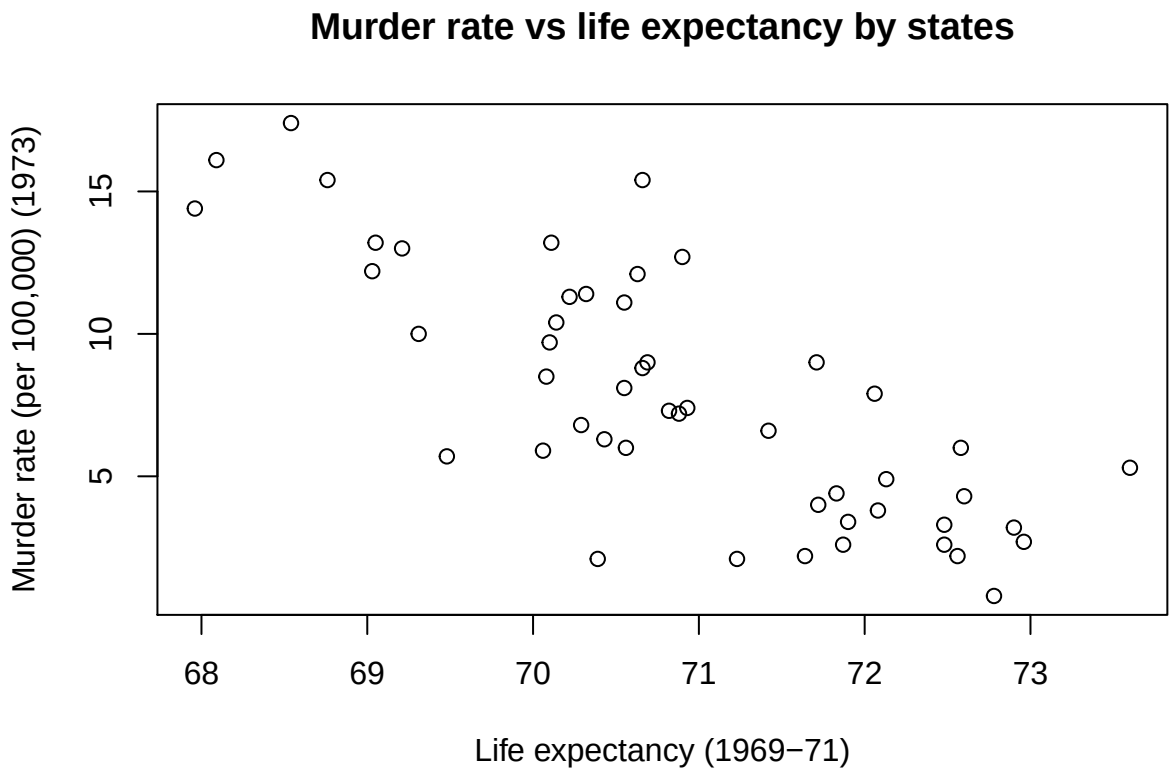
	ABB	Murder	Assault	UrbanPop	Rape	Population	Income	Illiteracy	Life.Exp	Murder
Connecticut	CT	3.3	110	77	11.1	3100	5348	1.1	72.48	3.1
Delaware	DE	5.9	238	72	15.8	579	4809	0.9	70.06	6.2
Florida	FL	15.4	335	80	31.9	8277	4815	1.3	70.66	10.7
Georgia	GA	17.4	211	60	25.8	4931	4091	2.0	68.54	13.9
Hawaii	HI	5.3	46	83	20.2	868	4963	1.9	73.60	6.2
Idaho	ID	2.6	120	54	14.2	813	4119	0.6	71.87	5.3
Illinois	IL	10.4	249	83	24.0	11197	5107	0.9	70.14	10.3
Indiana	IN	7.2	113	65	21.0	5313	4458	0.7	70.88	7.1
Iowa	IA	2.2	56	57	11.3	2861	4628	0.5	72.56	2.3
Kansas	KS	6.0	115	66	18.0	2280	4669	0.6	72.58	4.5
Kentucky	KY	9.7	109	52	16.3	3387	3712	1.6	70.10	10.6
Louisiana	LA	15.4	249	66	22.2	3806	3545	2.8	68.76	13.2
Maine	ME	2.1	83	51	7.8	1058	3694	0.7	70.39	2.7
Maryland	MD	11.3	300	67	27.8	4122	5299	0.9	70.22	8.5
Massachusetts	MA	4.4	149	85	16.3	5814	4755	1.1	71.83	3.3
Michigan	MI	12.1	255	74	35.1	9111	4751	0.9	70.63	11.1
Minnesota	MN	2.7	72	66	14.9	3921	4675	0.6	72.96	2.3
Mississippi	MS	16.1	259	44	17.1	2341	3098	2.4	68.09	12.5
Missouri	MO	9.0	178	70	28.2	4767	4254	0.8	70.69	9.3
Montana	MT	6.0	109	53	16.4	746	4347	0.6	70.56	5.0
Nebraska	NE	4.3	102	62	16.5	1544	4508	0.6	72.60	2.9
Nevada	NV	12.2	252	81	46.0	590	5149	0.5	69.03	11.5
New Hampshire	NH	2.1	57	56	9.5	812	4281	0.7	71.23	3.3
New Jersey	NJ	7.4	159	89	18.8	7333	5237	1.1	70.93	5.2
New Mexico	NM	11.4	285	70	32.1	1144	3601	2.2	70.32	9.7
New York	NY	11.1	254	86	26.1	18076	4903	1.4	70.55	10.9
North Carolina	NC	13.0	337	45	16.1	5441	3875	1.8	69.21	11.1
North Dakota	ND	0.8	45	44	7.3	637	5087	0.8	72.78	1.4
Ohio	OH	7.3	120	75	21.4	10735	4561	0.8	70.82	7.4
Oklahoma	OK	6.6	151	68	20.0	2715	3983	1.1	71.42	6.4
Oregon	OR	4.9	159	67	29.3	2284	4660	0.6	72.13	4.2
Pennsylvania	PA	6.3	106	72	14.9	11860	4449	1.0	70.43	6.1
Rhode Island	RI	3.4	174	87	8.3	931	4558	1.3	71.90	2.4
South Carolina	SC	14.4	279	48	22.5	2816	3635	2.3	67.96	11.6
South Dakota	SD	3.8	86	45	12.8	681	4167	0.5	72.08	1.7
Tennessee	TN	13.2	188	59	26.9	4173	3821	1.7	70.11	11.0
Texas	TX	12.7	201	80	25.5	12237	4188	2.2	70.90	12.2
Utah	UT	3.2	120	80	22.9	1203	4022	0.6	72.90	4.5
Vermont	VT	2.2	48	32	11.2	472	3907	0.6	71.64	5.5
Virginia	VA	8.5	156	63	20.7	4981	4701	1.4	70.08	9.5
Washington	WA	4.0	145	73	26.2	3559	4864	0.6	71.72	4.3
West Virginia	WV	5.7	81	39	9.3	1799	3617	1.4	69.48	6.7
Wisconsin	WI	2.6	53	66	10.8	4589	4468	0.7	72.48	3.0
Wyoming	WY	6.8	161	60	15.6	376	4566	0.6	70.29	6.9

Analysis on Viorent Crimes

Muder rate vs life expectancy

The scatter plot below shows the relation between murder rate and life expectancy.

```r
plot(UScombined$Life.Exp, UScombined$Murder,
     main = 'Murder rate vs life expectancy by states',
     xlab = 'Life expectancy (1969-71)', ylab = 'Murder rate (per 100,000) (1973)'
     )
```



The correlation test indicates that there are negative correlation of -0.7784985.

```r
cor.test(UScombined$Life.Exp, UScombined$Murder)
```

```
##
##  Pearson's product-moment correlation
##
## data:  UScombined$Life.Exp and UScombined$Murder
## t = -8.5934, df = 48, p-value = 2.836e-11
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.8686217 -0.6385121
## sample estimates:
##       cor
## -0.7784985
```

The result from a bivariate linear regression is shown below.

```r
lmr <- lm(UScombined$Murder~UScombined$Life.Exp)
summary(lmr)
```

```
##
## Call:
## lm(formula = UScombined$Murder ~ UScombined$Life.Exp)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.922 -1.582 -0.156  1.703  7.060
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         186.8206    20.8375   8.966 7.97e-12 ***
## UScombined$Life.Exp  -2.5259     0.2939  -8.593 2.84e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.762 on 48 degrees of freedom
## Multiple R-squared:  0.6061, Adjusted R-squared:  0.5979
## F-statistic: 73.85 on 1 and 48 DF,  p-value: 2.836e-11
```