A Statistical Analysis of Well-Being Indicators for Provinces in Turkey

Ayşe Rumeysa Muş and Şaban Dalaman

ISE 534 - Data Mining

Introduction and data set description

Every year Turkish Statistical Institute (TUIK) conducts surveys to compile, evaluate, analyse and publish statistics in the fields of economy, social issues, demography, culture, environment, science and technology, and in the other required areas.

In this work, we present a brief statistical analysis and interpretation of results for 2015 statistics related Provinces in Turkey. The dataset is provided by TUIK after completing their study for all provinces in Turkey.

Our goal in this project is to explore and show relations between features and outcomes such as level of happiness, hopefullness and life satisfaction. The data set includes 47 observations devided as 40 features and 7 responses for 81 provinces in Turkey.

The observations are listed below with subgroups:

- **Province**
- **Housing**
  - Number of rooms per person
  - Toilet presence per centage in dwellings
  - Percentage of house holds having problems with quality of dwellings
- **Work Life**
  - Employment rate
  - Unemployment rate
  - Average daily earnings
  - Job satisfaction rate
- **Income and wealth**
  - Savings deposit per capita
  - Percentage of house holds in middle or higher income groups
  - Percentage of house holds declaring to fail on meeting basic needs
- **Health**
  - Infant mortality rate
  - Life expectancy at birth
  - Number of applications per doctor
  - Satisfication rate with health status
  - Satisfication rate with public health services
- **Education**
  - Net schooling ratio of pre-primary education between the ages of 3and5
  - Average point of placement basic scores of the system for Transition to Secondary Education from Basic Education

  - Average points of the Transition to Higher Education Examination
  - Percentage of higher education graduates
  - Satisfaction rate with public education services

- **Environment**
  - Average of PM10 values of the stations (airpollution)
  - Forest area per km2
  - Percentage of population receiving waste services
  - Percentage of households having noise problems from the streets
  - Satisfaction rate with municipal cleaning services

- **Safety**
  - Murder rate (per million people)
  - Number of traffic accidents involving death or injury (per thousand people)
  - Percentage of people feeling safe when walking alone at night
  - Satisfication rate with public safety services

- **Civic engagement**
  - Voter turnout at local administrations
  - Rate of membership to political parties
  - Percentage of persons interested in union/association activities

- **Access to infrastructure services**
  - Number of internet subscriptions (per hundred persons)
  - Access rate of population to sewerage and pipesystem
  - Access rate to airport
  - Satisfaction rate with municipal public transport services

- **Social life**
  - Number of cinema and theatre audience (per hundred persons)
  - Shopping mall area per thousand people
  - Satisfication rate with social relations
  - Satisfication rate with social life

- **Key Responses**
  - Level of happiness
  - Hopeful
  - Life Satisfaction Index
  - Life Expectancy Total
  - Life Expectancy Male
  - Life Expectancy Female

Data pre-processing and statistical analysis of the data

The following chunks of codes contain preprocessing of the data set :

Reading the data from the file containing the data set.

Since data is already preprocessed by TUIK, there is no missing data and data cleaning is not required. There are two observation - "Access rate to airport" and "Shopping mall area per thousand people" that contains 0 entries. It seems they are showing actual situation for the related provinces. However we are going to investigate the side-effects of having 0 values in the model evaluation phase. There is no nominal attributes but there are attributes with different scales. In order to minimize bias for the models we are evaulating, we are going to use normalization techniques for some observed data.

Statistical Summary of some important features:

Top 5 correlated observations

```
## [1] "Life.Expectancy.Total vs."
## [1] "Life.expectancy.at.birth : 1.000000"
## [1] " "
## [1] "Life.Expectancy.Female vs."
## [1] "Life.expectancy.at.birth : 0.943800"
## [1] " "
## [1] "Access.rate.of.population.to.sewerage.and.pipesystem vs."
## [1] "Percentage.of.population.receiving.waste.services : 0.931386"
## [1] " "
## [1] "Life.Expectancy.Male vs."
## [1] "Life.expectancy.at.birth : 0.895234"
## [1] " "
## [1] "Satisfication.rate.with.public.safety.services vs."
## [1] "Satisfaction.rate.with.public.education.services : 0.890529"
## [1] " "
```

Top 5 observations covariances

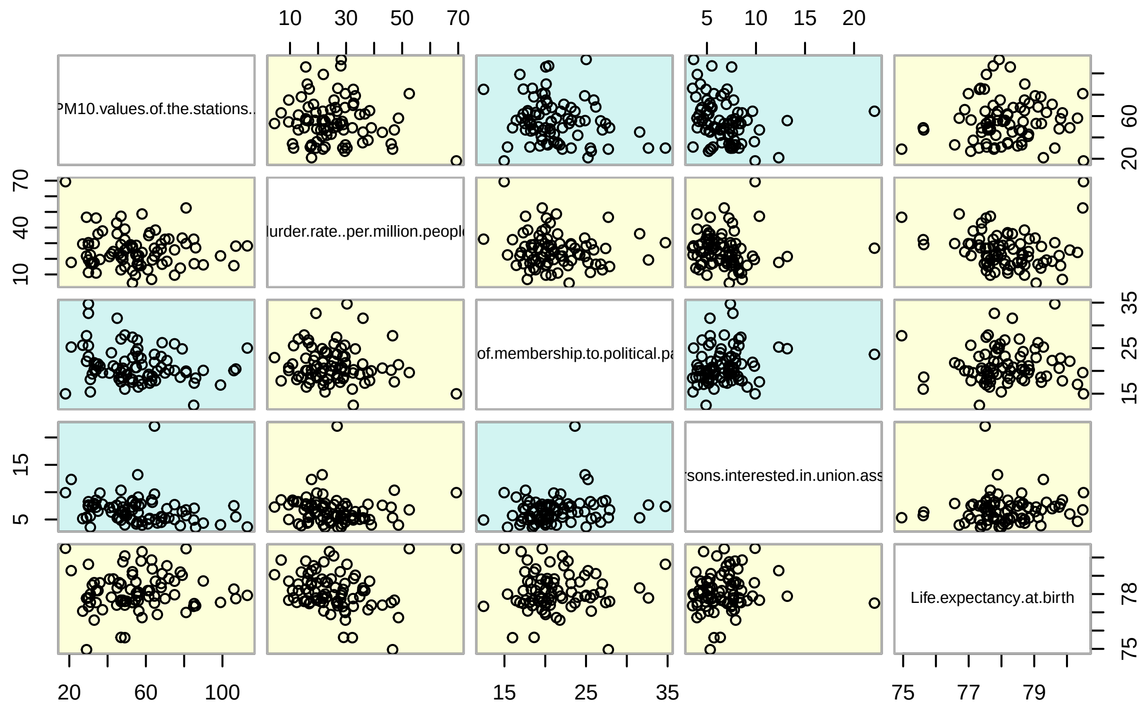
Before calculating covariances, min/max normalization was applied to observations values.

```
## [1] "Access.rate.of.population.to.sewerage.and.pipesystem vs."
## [1] "Percentage.of.population.receiving.waste.services : 0.053766"
## [1] " "
## [1] "Average.point.of.placement.basic.scores.of.the.system.for.Transition.to.Secondary.Education.from
## [1] "Number.of.rooms.per.person : 0.052070"
## [1] " "
## [1] "Percentage.of.house.holds.declaring.to.fail.on.meeting.basic.needs vs."
## [1] "Number.of.rooms.per.person : 0.051833"
## [1] " "
## [1] "Number.of.internet.subscriptions..per.hundred.persons. vs."
## [1] "Number.of.rooms.per.person : 0.048827"
## [1] " "
## [1] "Percentage.of.house.holds.having.problems.with.quality.of.dwellings vs."
## [1] "Number.of.rooms.per.person : 0.047886"
## [1] " "
```

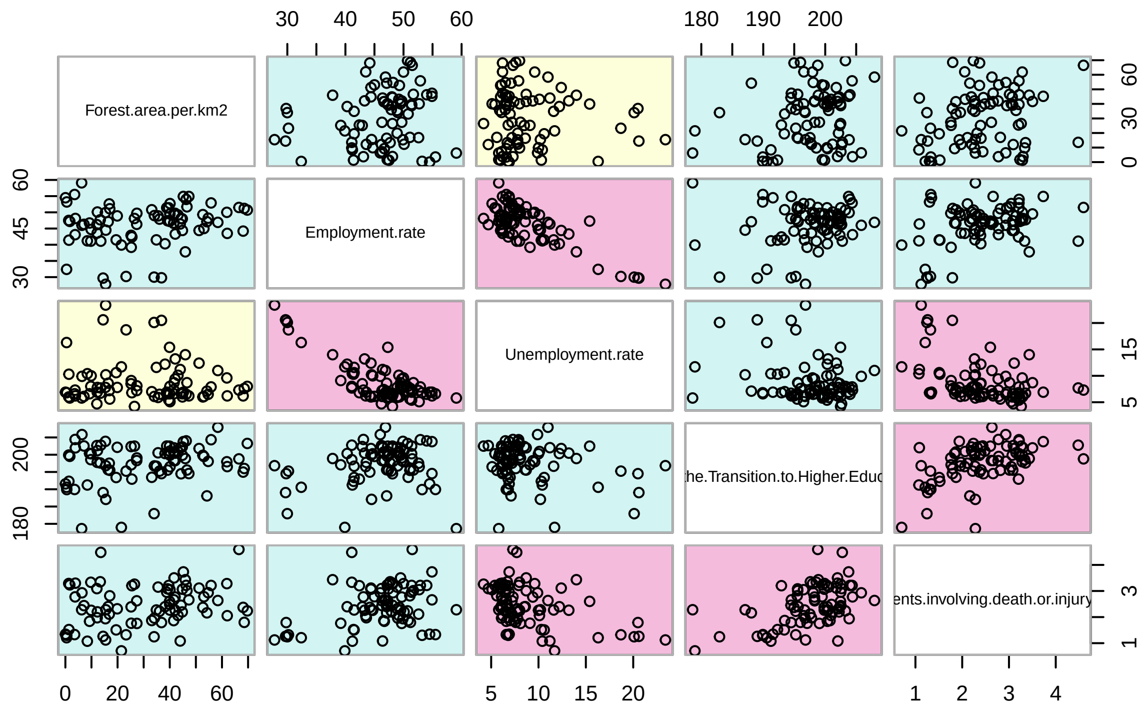
The grouping of observation correlations.

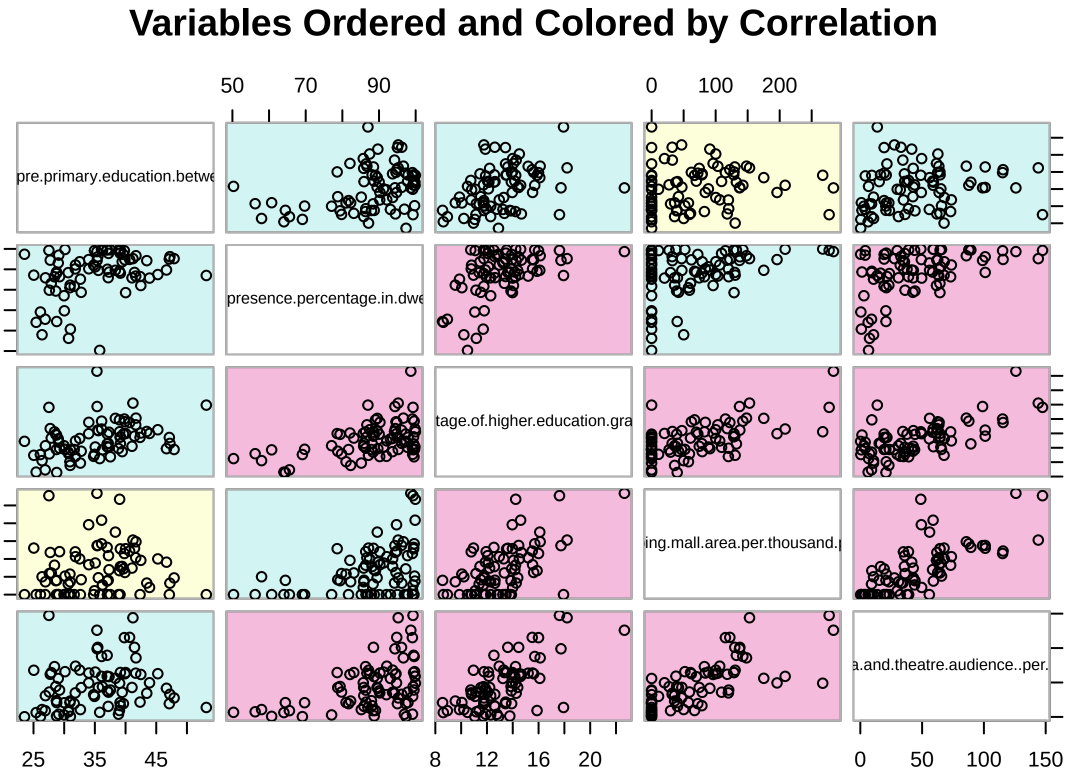
Each color shows different group.

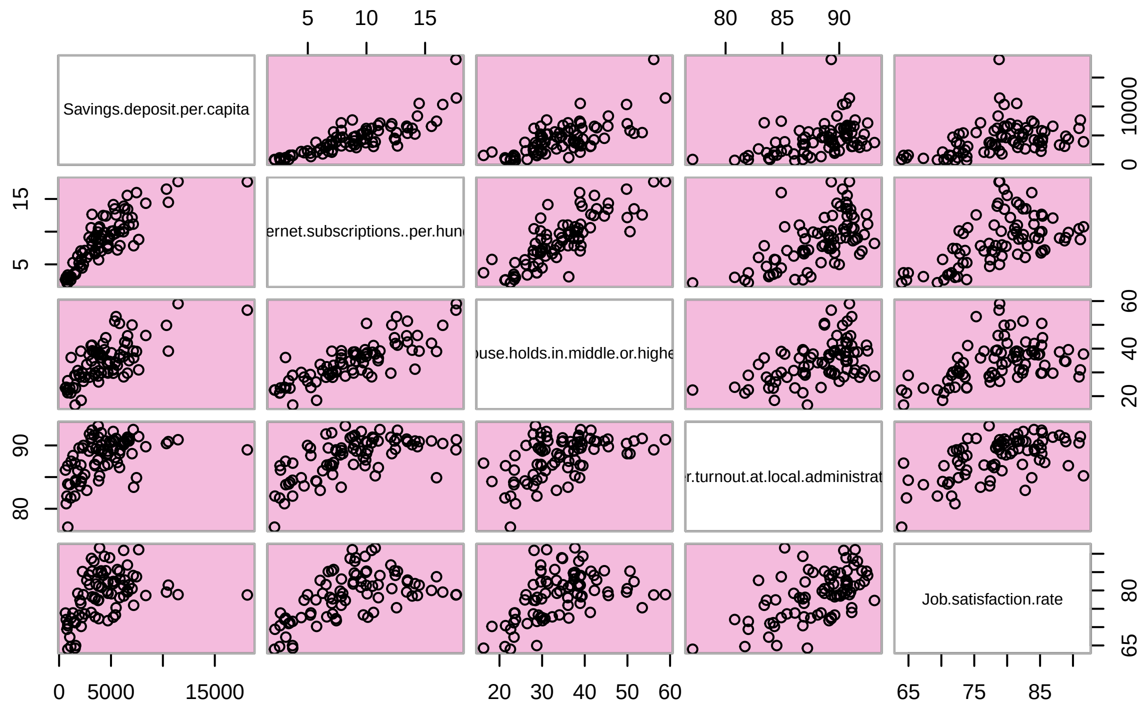
Variables Ordered and Colored by Correlation



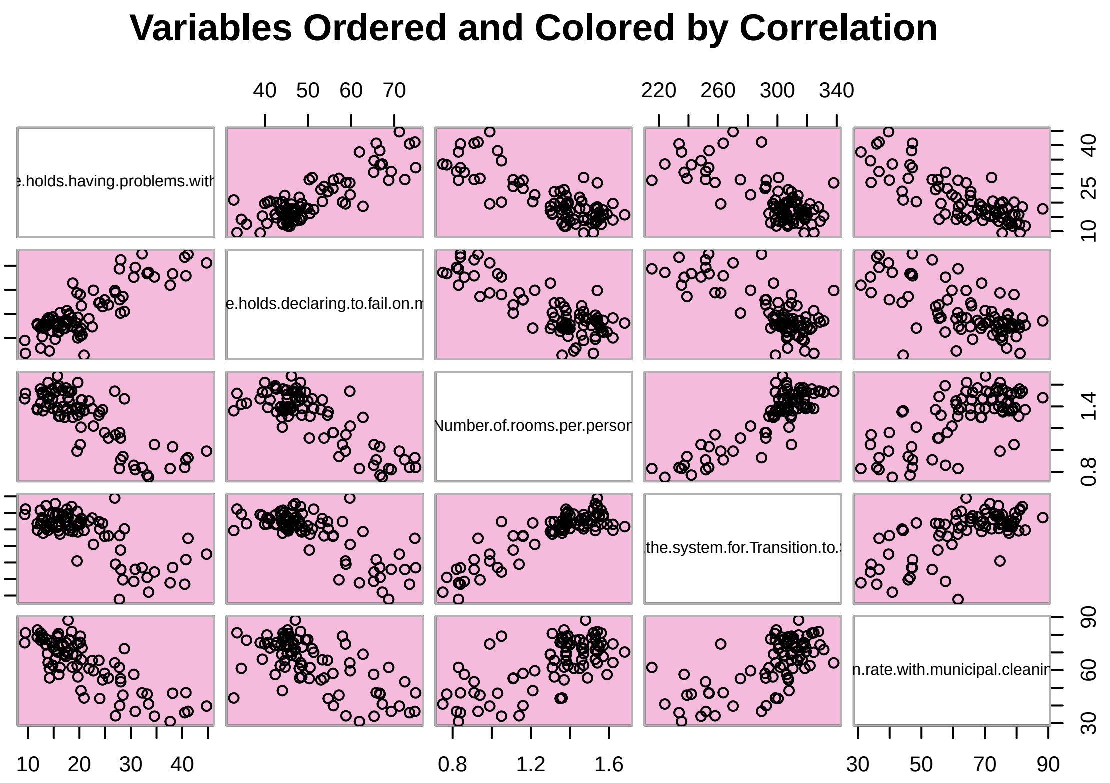
Variables Ordered and Colored by Correlation



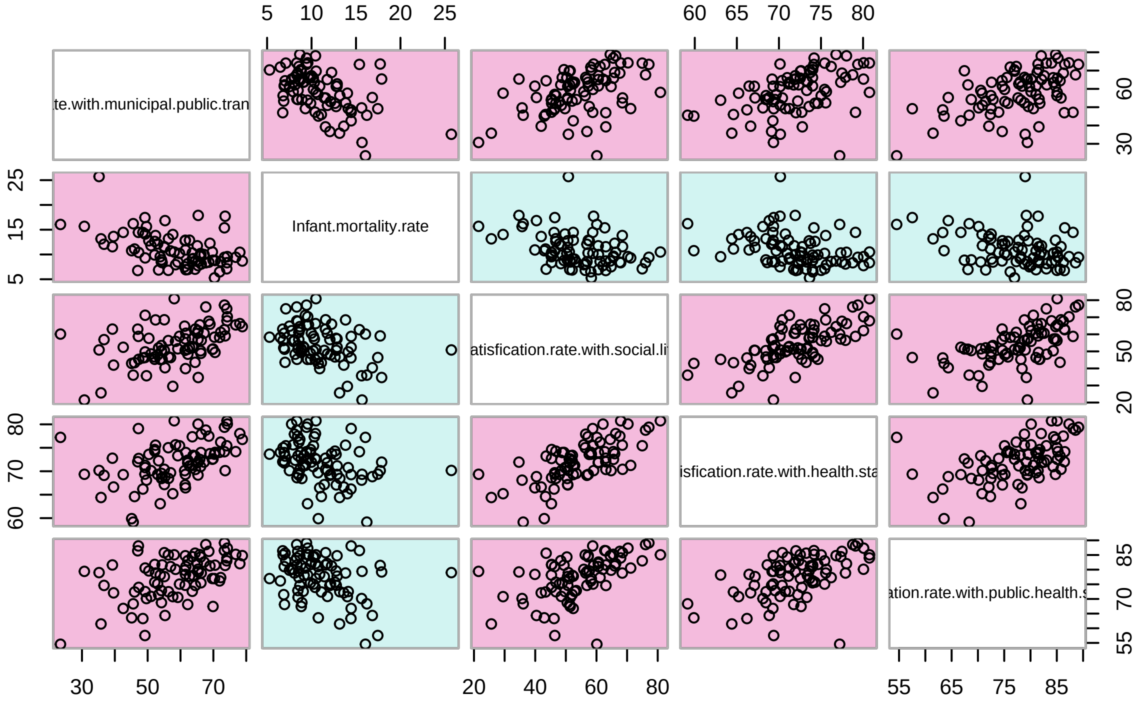

Variables Ordered and Colored by Correlation
pre.primary.education.betwe
presence.percentage.in.dwe
tage.of.higher.education.gra
ing.mall.area.per.thousand.p
a.and.theatre.audience..per.
50
70
90
0
100
200
25
40
14
20
8
0
150
80
50
100
25
35
45
8
12
16
20
0
50
100
150

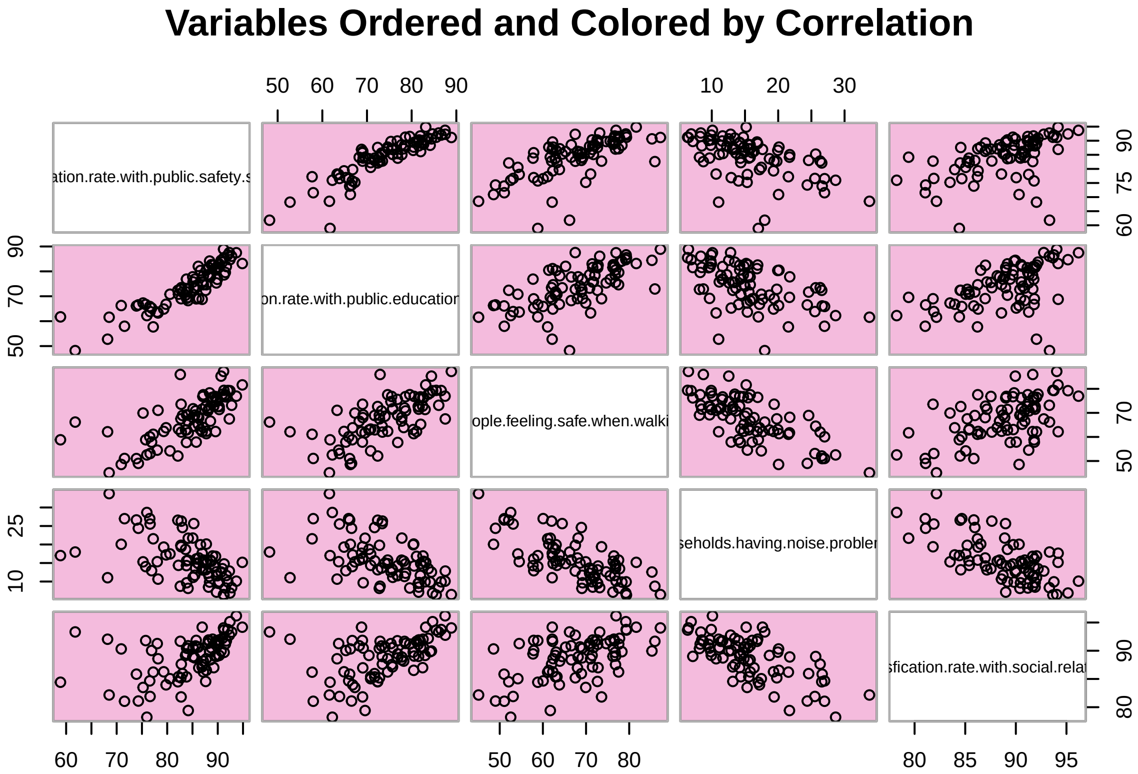


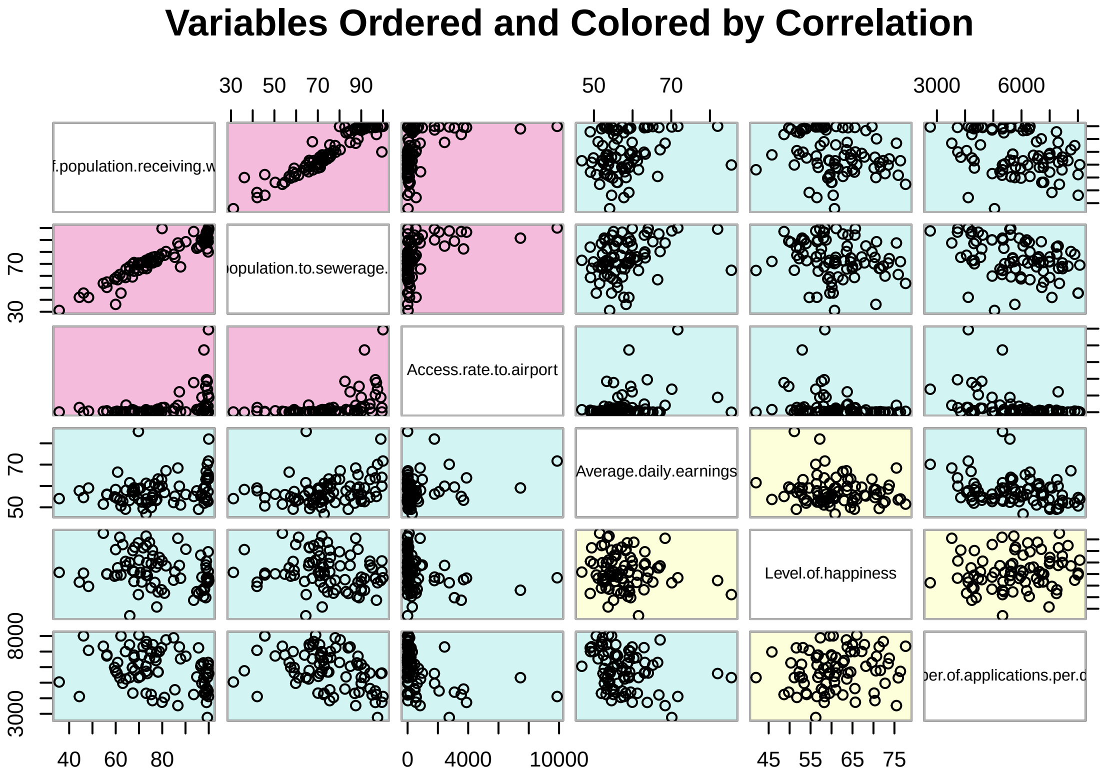
Variables Ordered and Colored by Correlation


Variables Ordered and Colored by Correlation
e.holds.having.problems.with
e.holds.declaring.to.fail.on.m
Number.of.rooms.per.person
the.system.for.Transition.to.S
n.rate.with.municipal.cleanin
40
50
60
70
220
260
300
340
40
25
10
60
40
1.4
0.8
300
220
90
60
30
10
20
30
40
0.8
1.2
1.6
30
50
70
90

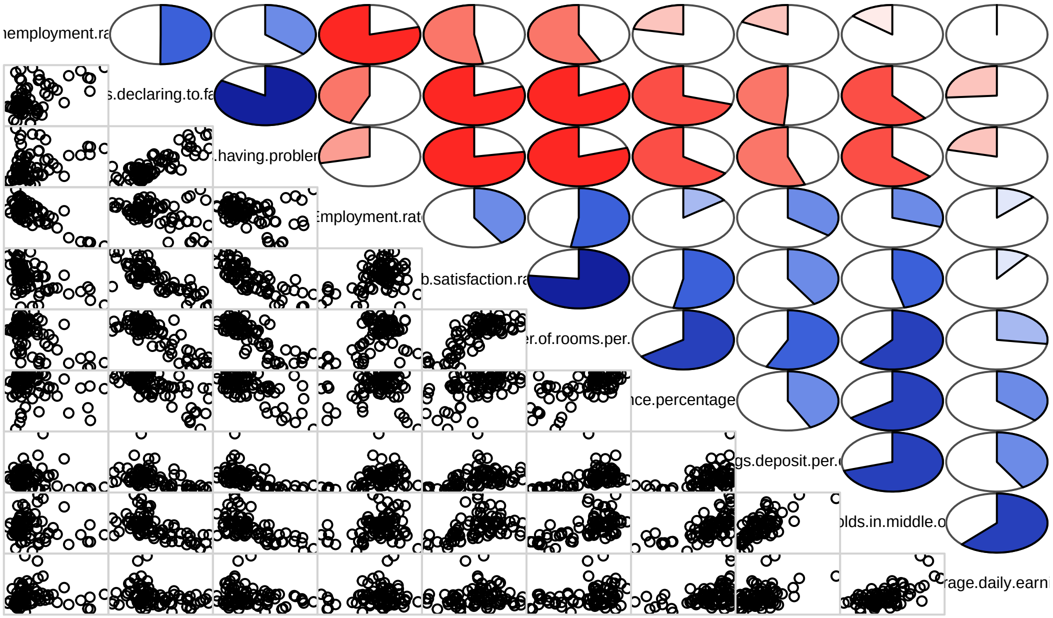
Variables Ordered and Colored by Correlation



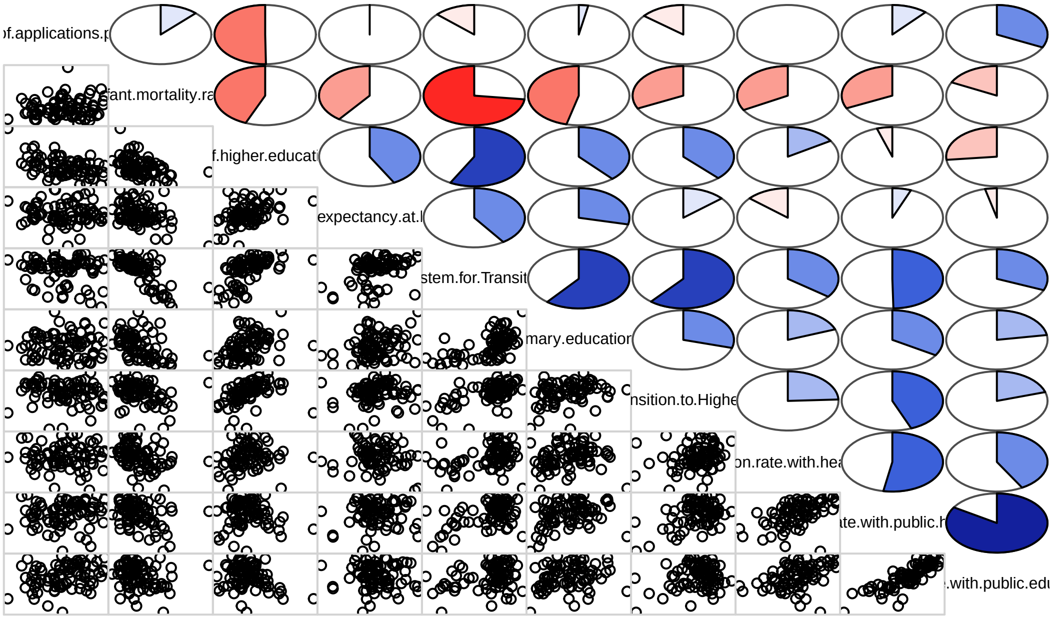

Variables Ordered and Colored by Correlation
ation.rate.with.public.safety.s
on.rate.with.public.education
ople.feeling.safe.when.walki
seholds.having.noise.probler
sfication.rate.with.social.rela
50
60
70
80
90
10
20
30
60
75
90
50
70
90
70
50
25
10
90
80
60
70
80
90
50
60
70
80
80
85
90
95


Variables Ordered and Colored by Correlation
f.population.receiving.w
opulation.to.sewerage.
Access.rate.to.airport
Average.daily.earnings
Level.of.happiness
ber.of.applications.per.d
30
50
70
90
50
70
3000
6000
80
40
70
30
6000
0
70
50
65
45
8000
3000
40
60
80
0
4000
10000
45
55
65
75

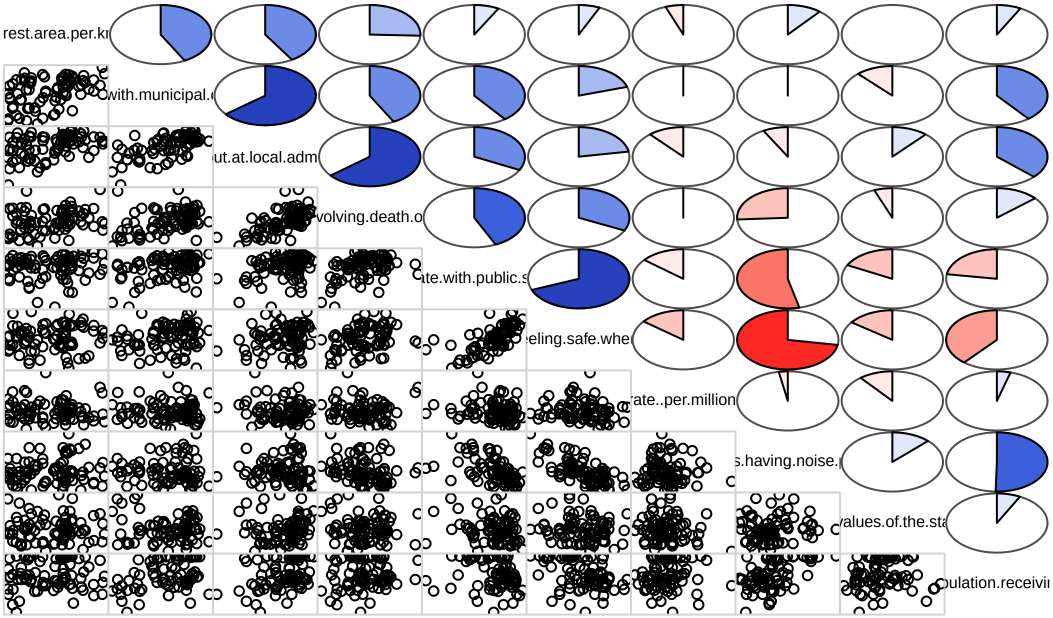
Variables Colored and Showed by Correlation Magnitude



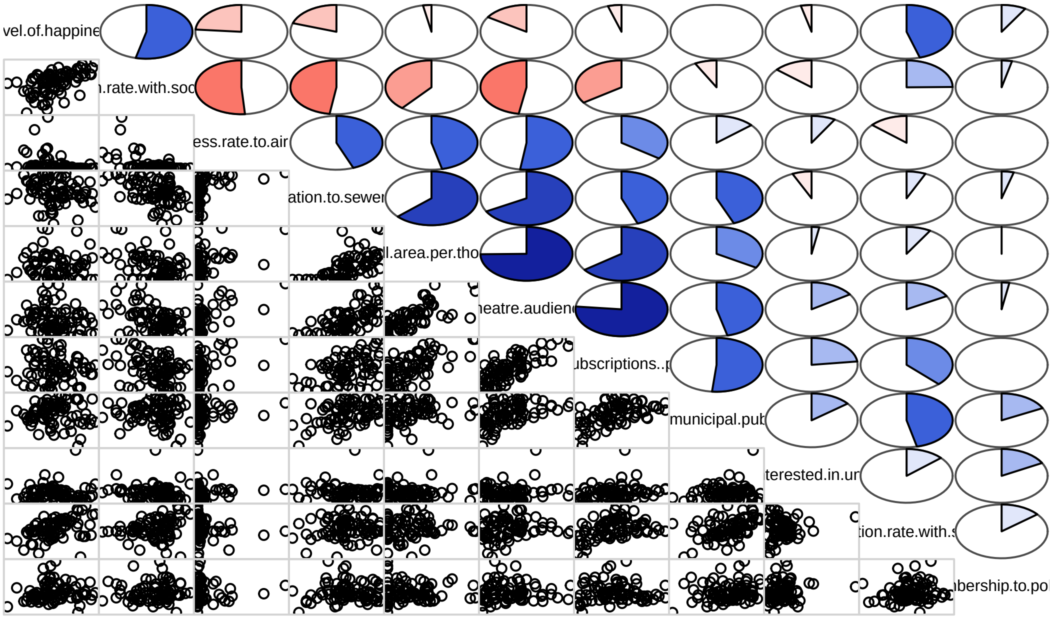
Variables Colored and Showed by Correlation Magnitude



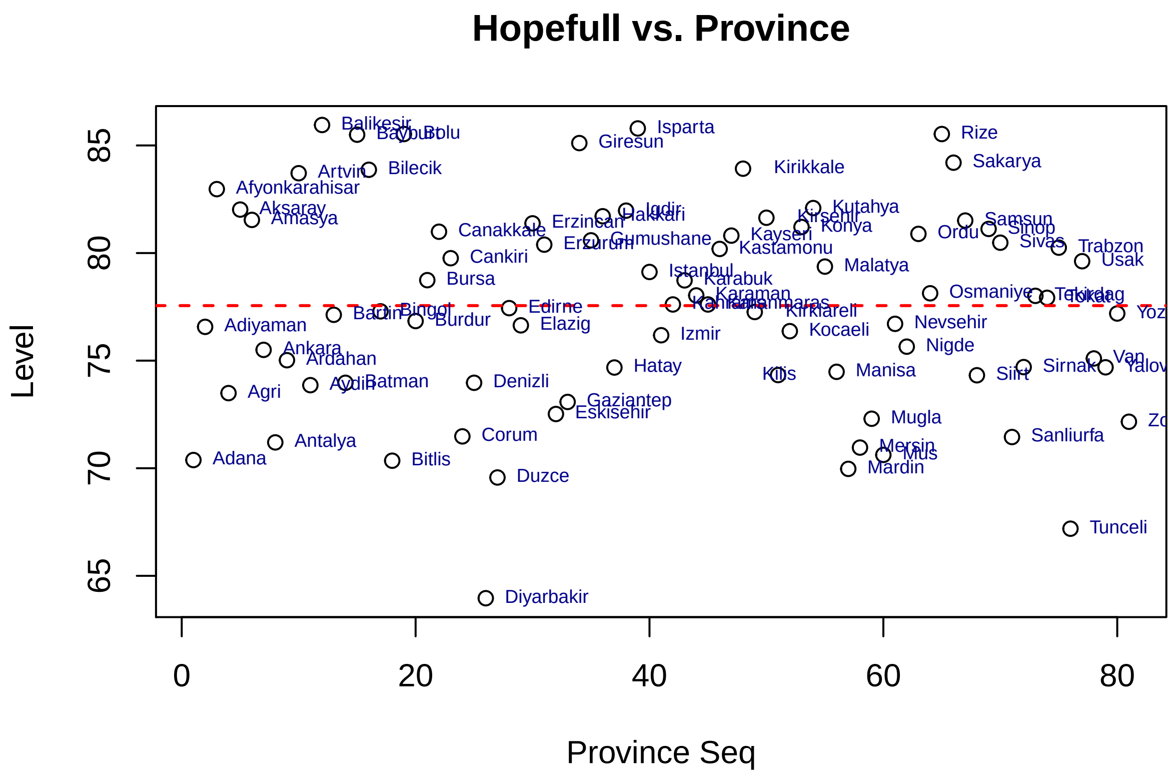
Variables Colored and Showed by Correlation Magnitude

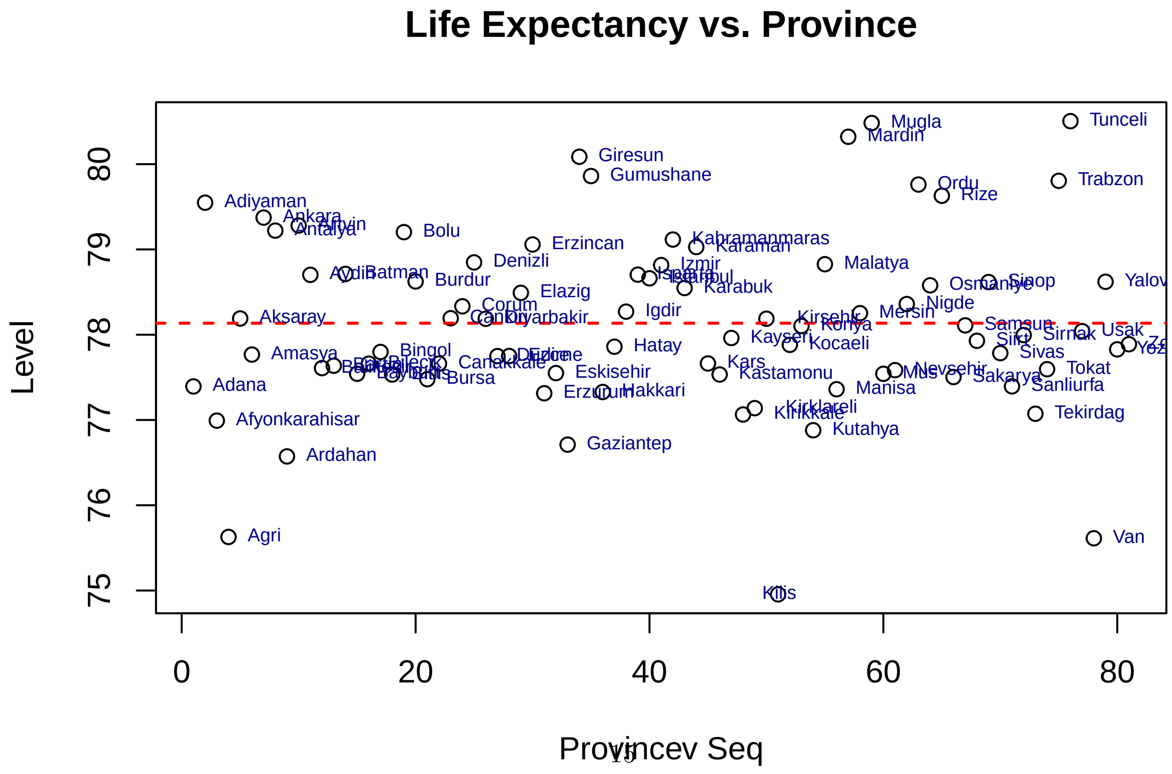


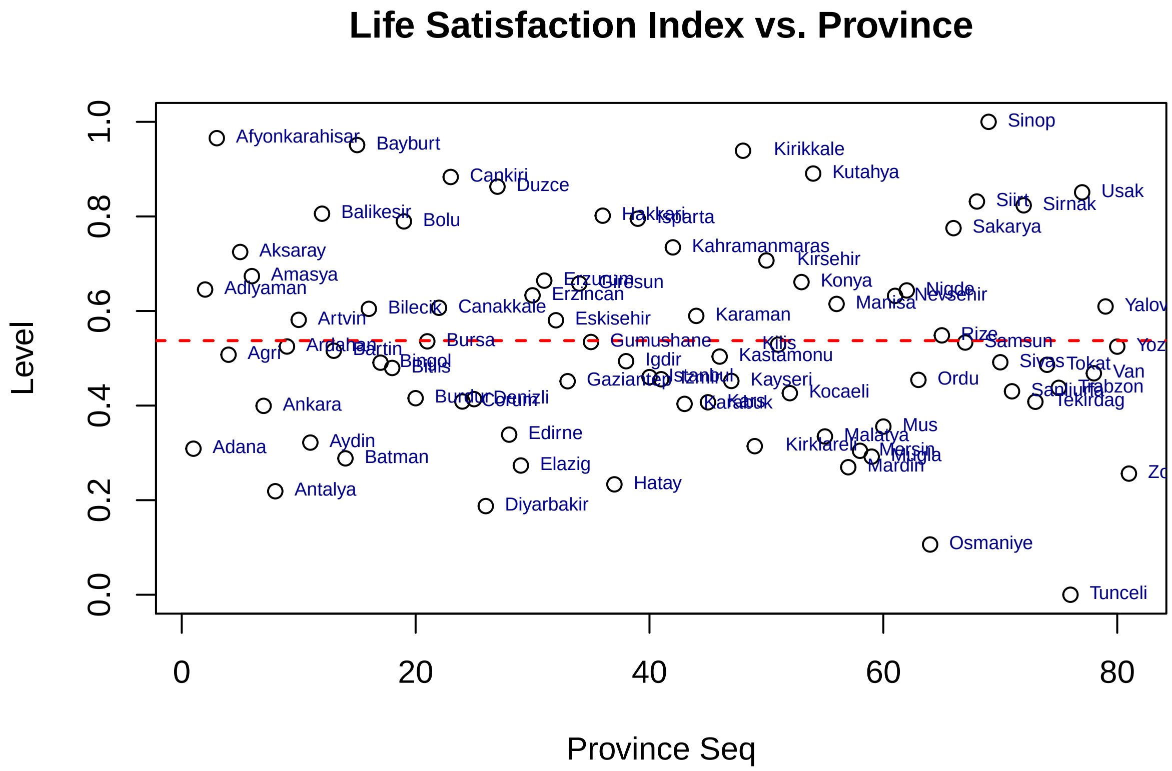
Variables Colored and Showed by Correlation Magnitude

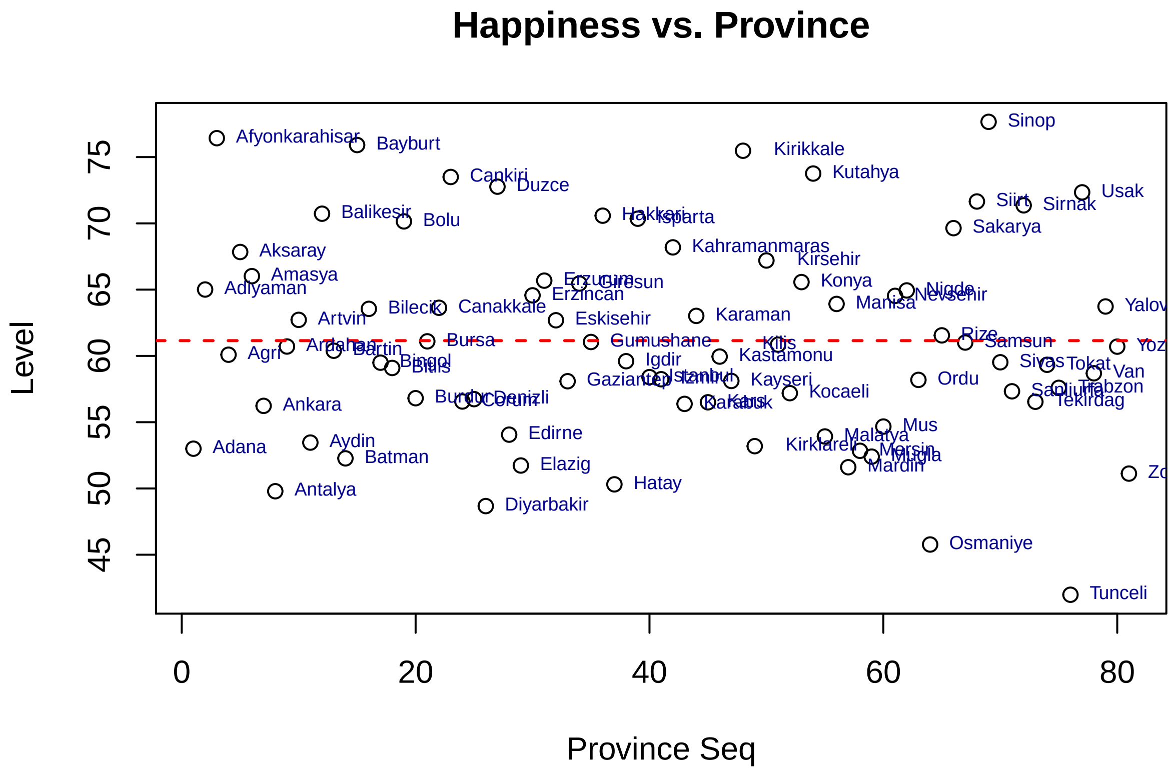


Hopefulness and Life satisfaction observation scores for each province.

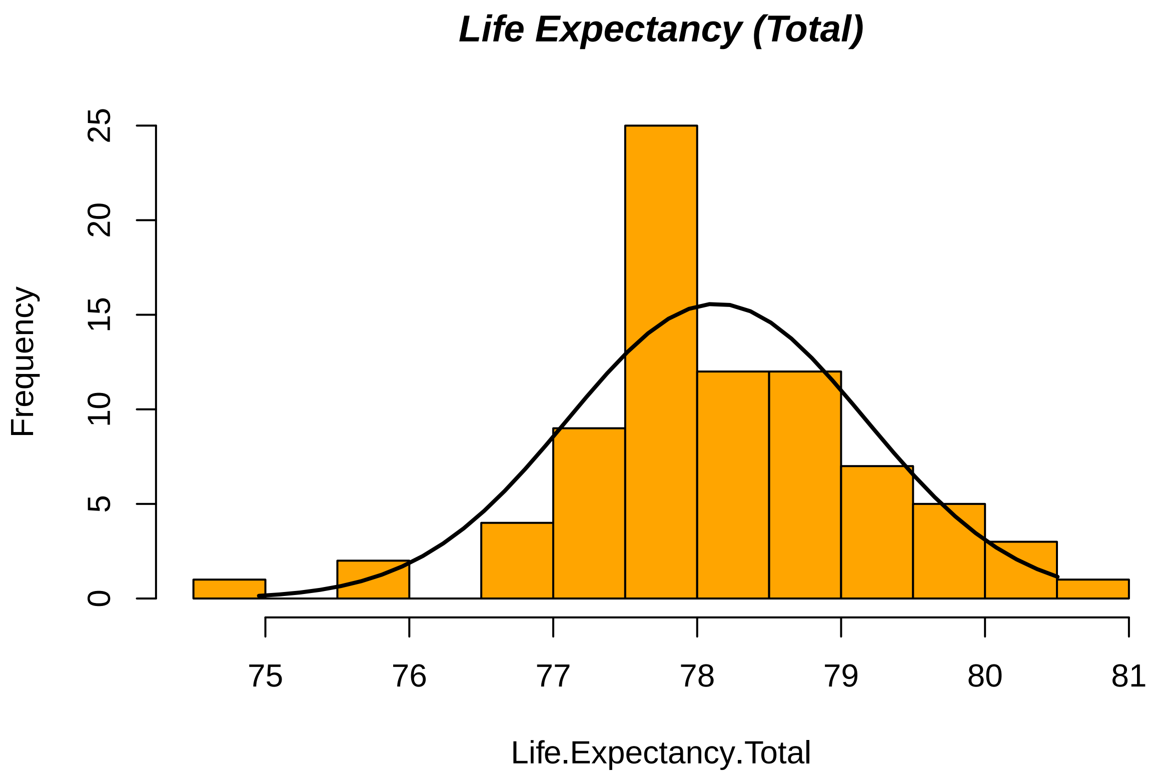


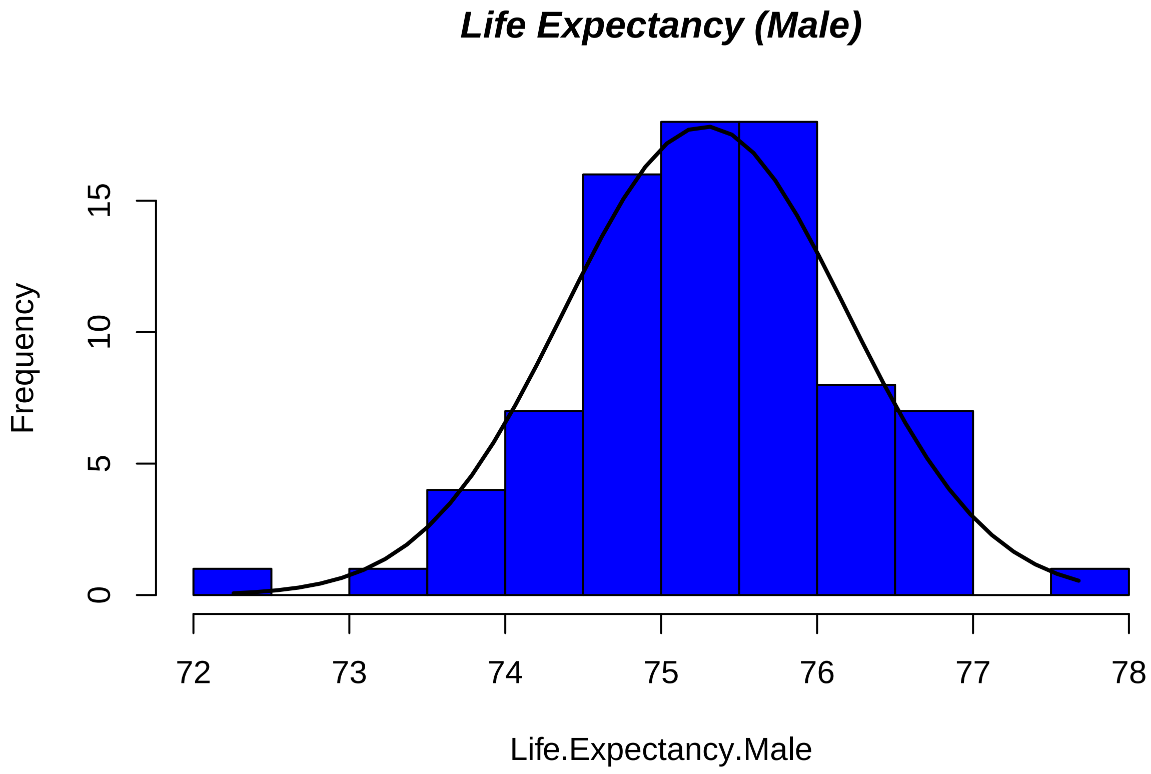


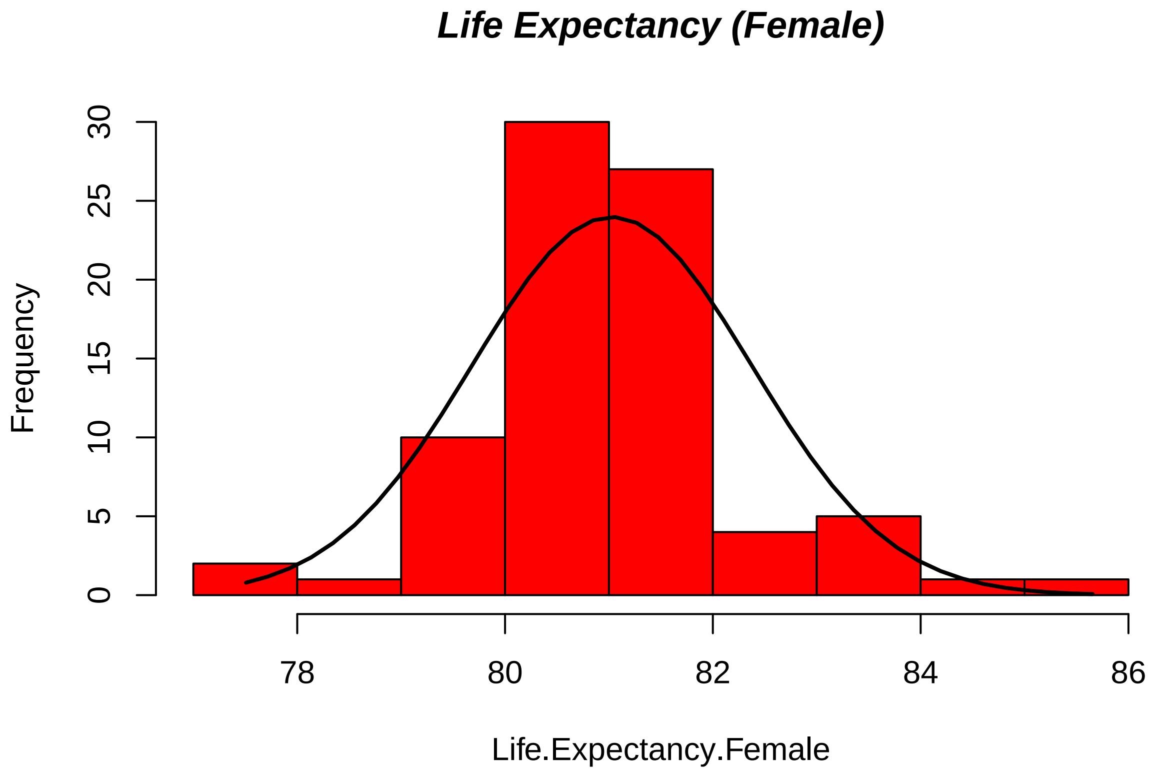

Life Satisfaction Index vs. Province
Level
Province Seq
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
Afyonkarahisar
Bayburt
Cankiri
Duzce
Balikesir
Bolu
Aksaray
Amasya
Adiyaman
Artvin
Bilecik
Canakkale
Agri
Ardahan
Bartin
Bursa
Bingol
Bitlis
Ankara
Burdur
Denizli
Corum
Adana
Aydin
Batman
Antalya
Edirne
Elazig
Diyarbakir
Hakkari
Isparta
Kahramanmaras
Erzurum
Giresun
Erzincan
Eskisehir
Gumushane
Igdir
Gaziantep
Istanbul
Izmir
Hatay
Kirikkale
Kutahya
Kirsehir
Konya
Manisa
Nigde
Nevsehir
Karaman
Kilis
Kastamonu
Kayseri
Karabuk
Kars
Kocaeli
Kirklareli
Malatya
Mus
Mersin
Mugla
Mardin
Ordu
Osmaniye
Sinop
Siirt
Sirnak
Usak
Sakarya
Yalov
Rize
Samsun
Yoz
Sivas
Tokat
Van
Sanliurfa
Trabzon
Tekirdag
Zo
Tunceli


Happiness vs. Province
Level
Province Seq
75
70
65
60
55
50
45
0
20
40
60
80
Sinop
Afyonkarahisar
Bayburt
Kirikkale
Kutahya
Cankiri
Duzce
Usak
Siirt
Sirnak
Balikesir
Bolu
Hakkari
Isparta
Sakarya
Kahramanmaras
Aksaray
Kirsehir
Amasya
Adiyaman
Erzurum
Giresun
Erzincan
Konya
Nigde
Nevsehir
Manisa
Bilecik
Canakkale
Yalov
Artvin
Eskisehir
Karaman
Rize
Ardahan
Bartin
Bursa
Gumushane
Kilis
Samsun
Yoz
Agri
Kastamonu
Bingol
Bitlis
Igdir
Sivas
Tokat
Van
Gaziantep
Istanbul
Izmir
Kayseri
Ordu
Kocaeli
Sanliurfa
Trabzon
Burdur
Denizli
Corum
Karabuk
Kars
Tekirdag
Ankara
Mus
Edirne
Malatya
Kirklareli
Aydin
Adana
Mersin
Mugla
Batman
Mardin
Elazig
Zo
Hatay
Antalya
Diyarbakir
Osmaniye
Tunceli

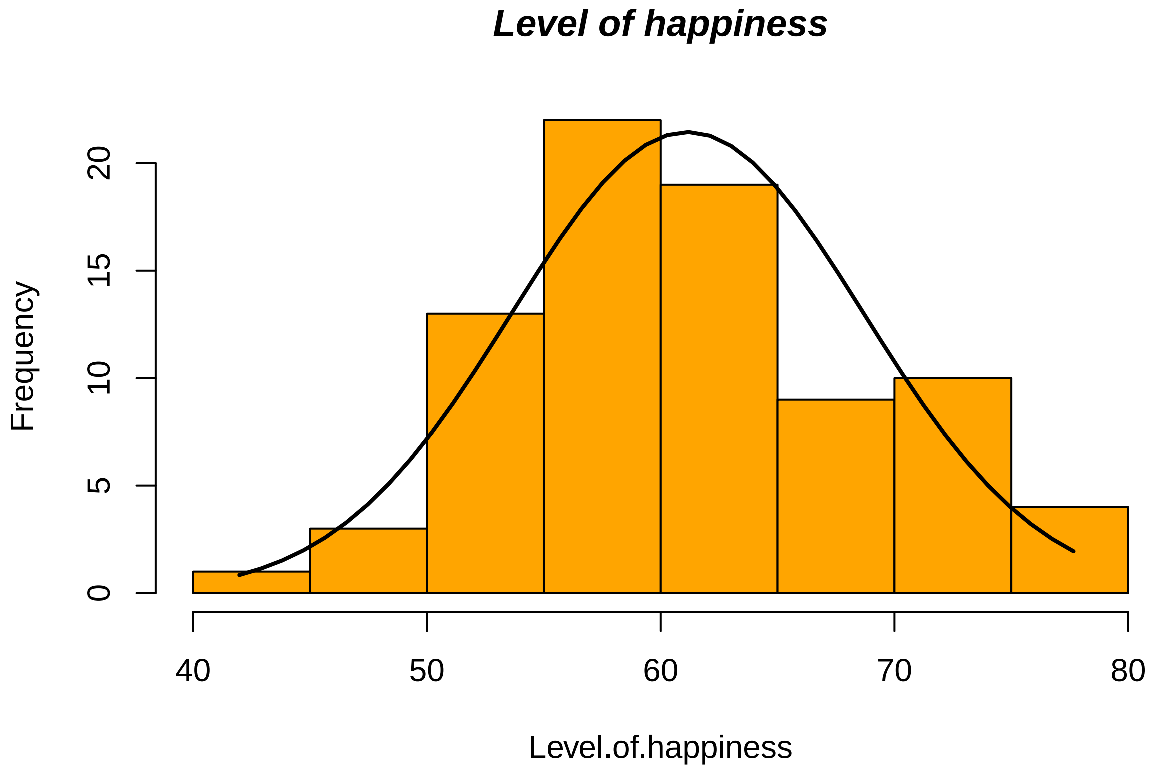
The distribution of Life Expectancy attributes for all provinces.

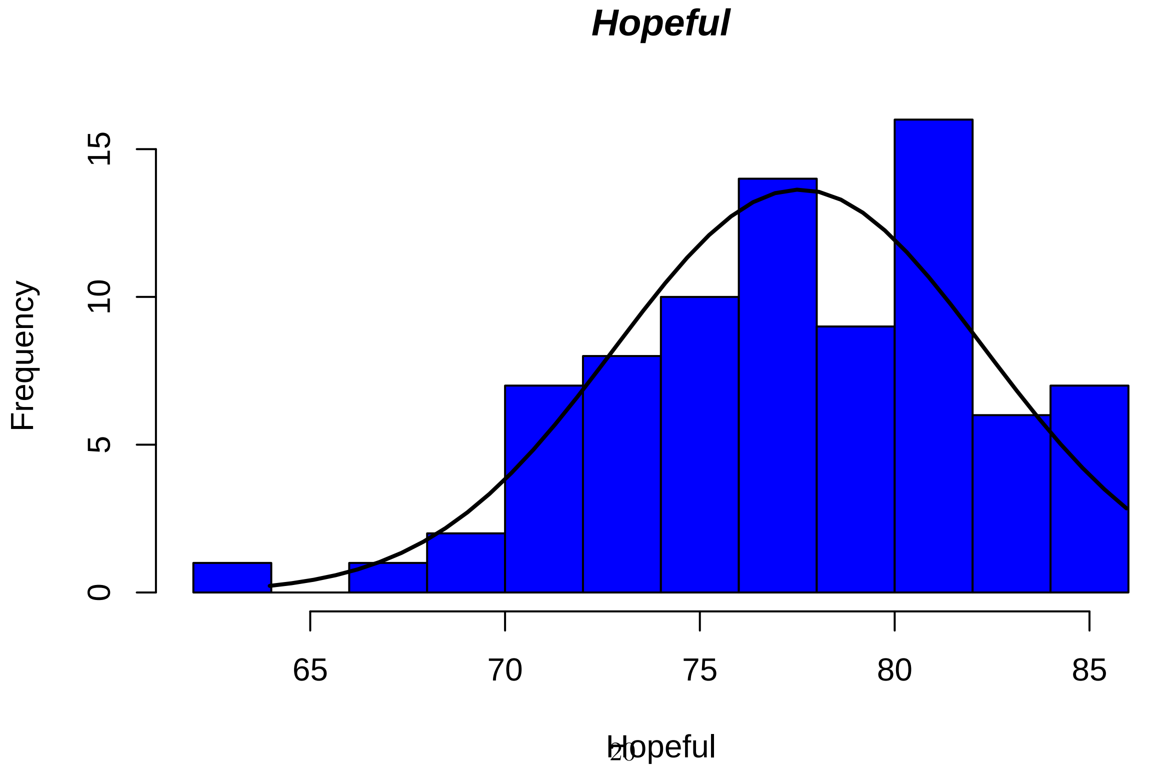


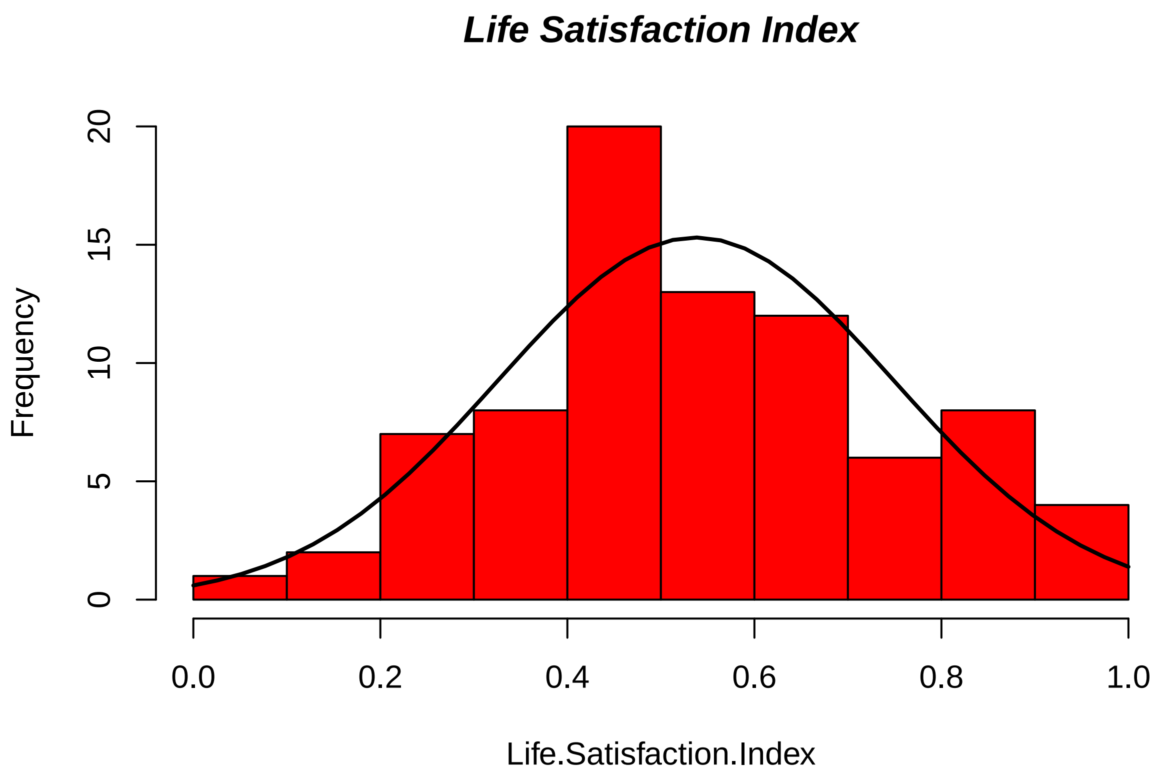



Life Expectancy (Female)
Frequency
0
5
10
15
20
25
30
78
80
82
84
86
Life.Expectancy.Female

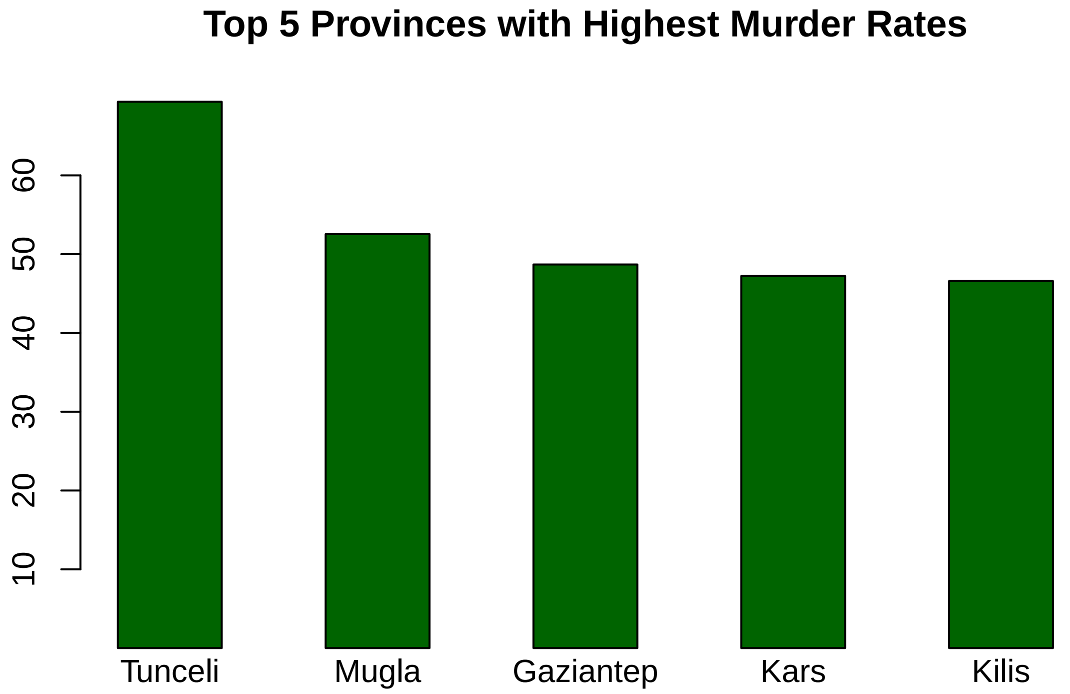
The distribution of Level of Happiness and Hope attributes for all provinces.

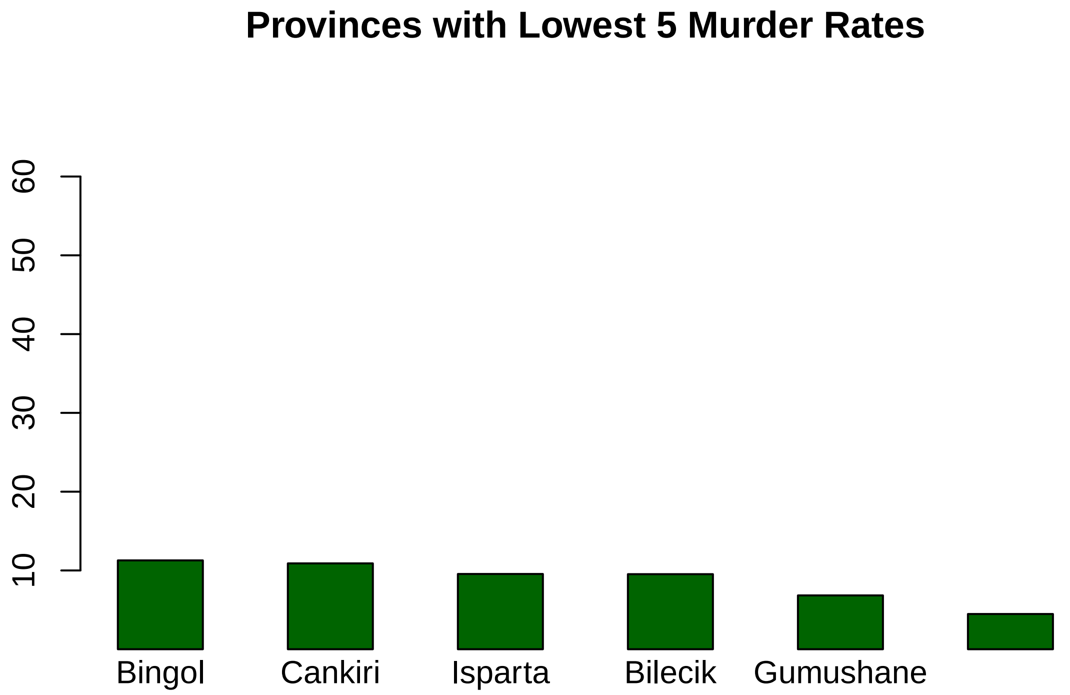





Life Satisfaction Index
Frequency
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
Life.Satisfaction.Index

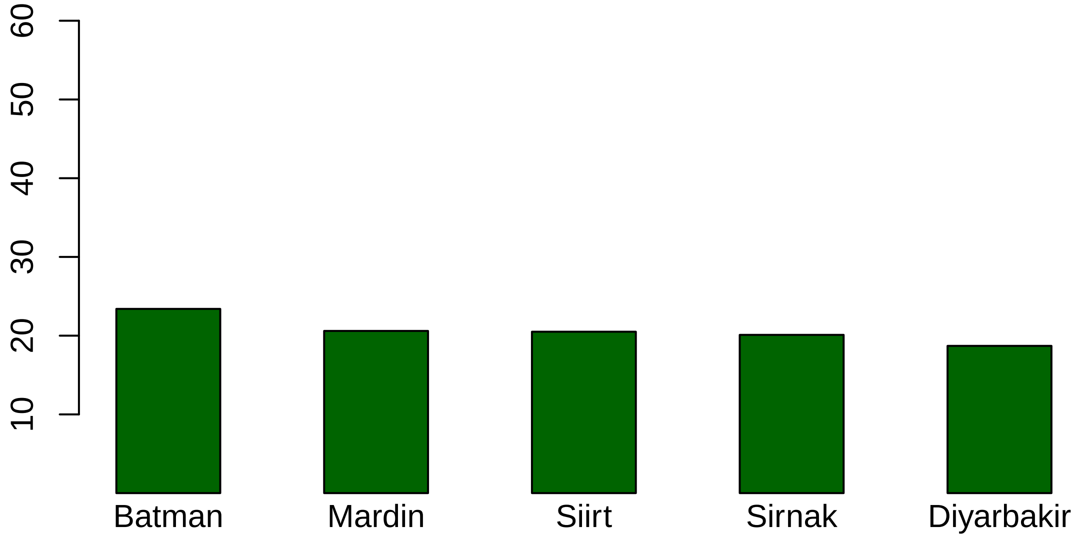
Top 5 and lowest 5 Murder rates among all provinces


Top 5 Provinces with Highest Murder Rates

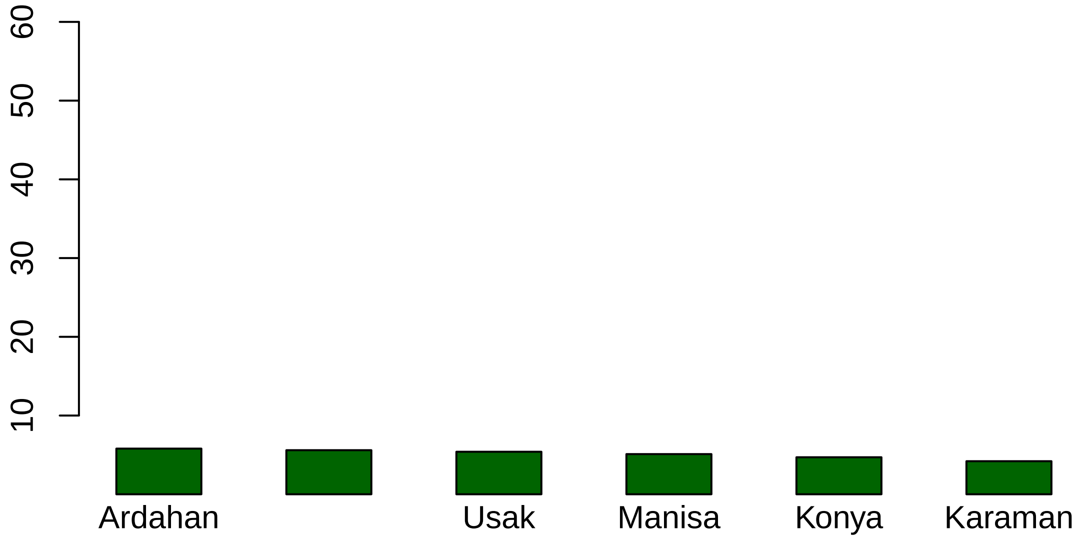

Provinces with Lowest 5 Murder Rates

Top 5 and lowest 5 Unemployment rates among all provinces

Top 5 Provinces with Highest Unemployment rate



Provinces with Lowest 5 Unemployment rate


Ardahan
Usak
Manisa
Konya
Karaman