

Unit 10 activities

Guilherme Amorim

2024-10-16

Unit 10 data activities

1. Find out correlation between systolic and diastolic BP.
2. Produce a scatter plot between systolic and diastolic BP.

R

Initial setup

```
# initial setup
library(haven)
library(skimr)
library(tidyverse)
library(here)
```

```
## Warning: package 'here' was built under R version 4.3.3
```

Data load

```
Health_Data<-read_sav(here("Datasets/Health Data.sav"))
```

Correlation

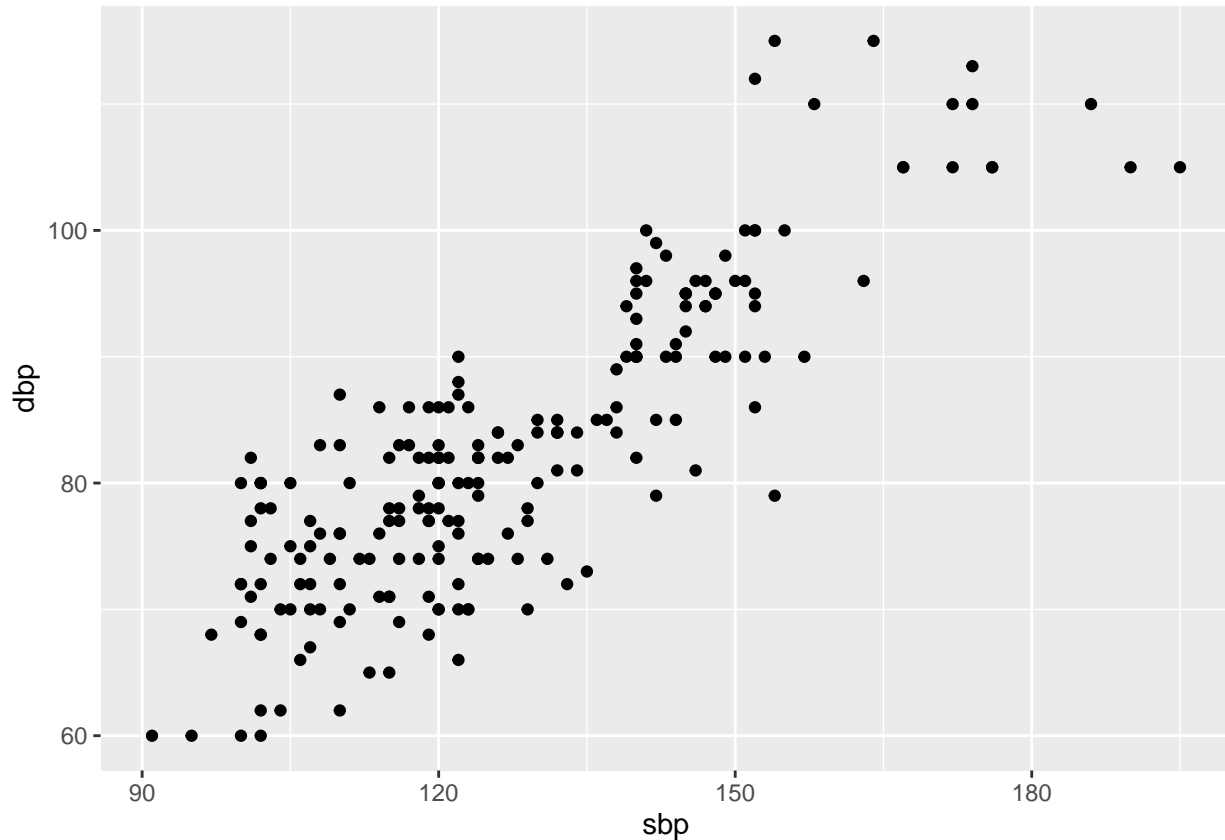
```
cor.test(Health_Data$sbp, Health_Data$dbp)

##
## Pearson's product-moment correlation
##
## data: Health_Data$sbp and Health_Data$dbp
## t = 22.961, df = 208, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.8035436 0.8811723
## sample estimates:
## cor
## 0.846808
```

r (Pearson's correlation coefficient) = 0.846, p -value < 0.001, indicating a statistically significant strong correlation between systolic and diastolic blood pressure

Scatterplot

```
Health_Data%>%  
  ggplot(aes(sbp, dbp))+  
  geom_point()
```



Python

Initial setup

```
# first install packages (from terminal)  
  
# pip3 install numpy  
  
# pip3 install pandas  
  
# pip3 install matplotlib  
  
# pip3 install pyreadstat
```

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from scipy import stats
```

Data load

```
Health_Data_python = pd.read_spss("C:/Users/guilhermep/Documents/PgDip/Coding/Module 2/pgdip_module2_pr
```

Correlation

```
stats.pearsonr(Health_Data_python.sbp, Health_Data_python.dbp)
```

```
## PearsonRResult(statistic=0.8468079888445823, pvalue=6.1123324179697985e-59)
```

Scatterplot

```
plt.scatter(Health_Data_python.sbp, Health_Data_python.dbp, alpha=0.5)
plt.show()
```

