"""

Name:

Course:

Project Name:

"""

1. Data loading

Sex Seat libarts TV ... momheight dadheight exercise GPA

0 Female Back NonLib 13.0 ... 66.0 71.0 10.0 4.00

1 Male Back NonLib 20.0 ... 64.0 65.0 2.0 2.30

2 Male Back NonLib 15.0 ... 62.0 74.0 3.0 2.60

3 Male Middle NonLib 8.0 ... 59.0 70.0 6.0 2.80

4 Female Front NonLib 2.5 ... 65.0 70.0 6.5 2.62

.. ... ... ... ... ... ... ... ... ...

150 Female Back NonLib 5.0 ... 69.0 72.0 5.0 2.18

151 Male Back NonLib 8.0 ... 65.0 71.0 1.5 2.80

152 Male Middle NonLib 9.0 ... 68.0 74.0 7.0 3.50

153 Male Front NonLib 1.0 ... 60.0 74.0 3.0 4.00

154 Female Middle NonLib 5.0 ... 62.0 66.0 6.0 4.00

[155 rows x 12 columns]

Central tendancy

Mean

TV 8.741935

computer 14.068182

Sleep 6.985484

alcohol 4.409091

Height 66.893548

momheight 63.288710

dadheight 69.256452

exercise 4.597403

GPA 2.913582

dtype: float64

Median

TV 6.0

computer 10.0

Sleep 7.0

alcohol 1.0

Height 66.5

momheight 63.0

dadheight 69.0

exercise 3.0

GPA 3.0

dtype: float64

Mode

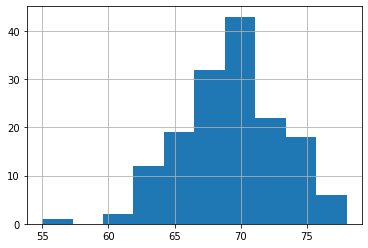
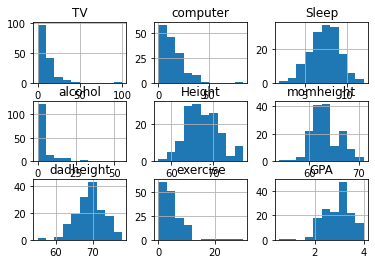
Sex Seat libarts TV ... momheight dadheight exercise GPA

0 Female Middle NonLib 1.0 ... 62.0 68.0 2.0 3.0

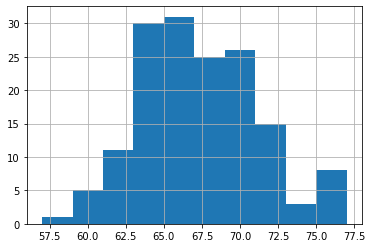
[1 rows x 12 columns]

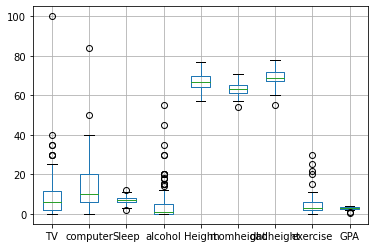
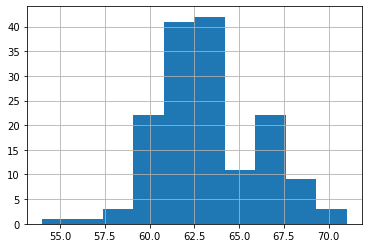
2. histogram

All data histogram

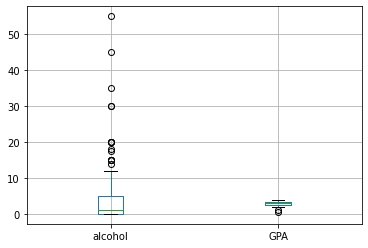
Dad height

Height

Mom height



All data box plot

Alcohol GPA box plot

3. They indicate how much the observations are spread out around, let’s say, “a center”.

4.

66.0 9.677419

65.0 9.032258

64.0 9.032258

63.0 8.387097

68.0 7.741935

70.0 7.741935

69.0 7.741935

67.0 7.096774

72.0 5.806452

61.0 3.870968

71.0 3.225806

62.0 2.580645

60.0 2.580645

75.0 2.580645

73.0 1.935484

76.0 1.290323

77.0 1.290323

63.5 1.290323

66.5 0.645161

72.5 0.645161

67.5 0.645161

57.0 0.645161

68.5 0.645161

69.5 0.645161

59.0 0.645161

62.5 0.645161

65.5 0.645161

70.5 0.645161

64.5 0.645161

Name: Height, dtype: float64

62.00 17.419355

64.00 16.129032

60.00 14.193548

63.00 10.322581

61.00 9.032258

66.00 8.387097

65.00 7.096774

67.00 5.806452

68.00 3.225806

69.00 1.935484

70.00 1.290323

59.00 1.290323

63.50 0.645161

57.00 0.645161

71.00 0.645161

58.00 0.645161

69.25 0.645161

54.00 0.645161

Name: momheight, dtype: float64

68.00 12.258065

70.00 11.612903

72.00 9.677419

74.00 9.032258

69.00 9.032258

67.00 7.741935

71.00 7.096774

65.00 6.451613

66.00 5.806452

73.00 3.870968

64.00 3.870968

63.00 2.580645

76.00 1.935484

75.00 1.935484

62.00 1.290323

77.00 1.290323

66.50 0.645161

71.50 0.645161

78.00 0.645161

55.00 0.645161

74.75 0.645161

61.00 0.645161

60.00 0.645161

Name: dadheight, dtype: float64

Height dadheight momheight

69.0 68.0 65.0 1.290323

64.0 68.0 62.0 1.290323

63.0 70.0 61.0 1.290323

57.0 66.0 61.0 0.645161

68.0 74.0 62.0 0.645161

65.0 69.0 63.5 0.645161

70.0 64.0 0.645161

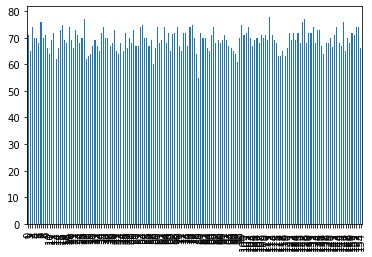
72.0 64.0 0.645161

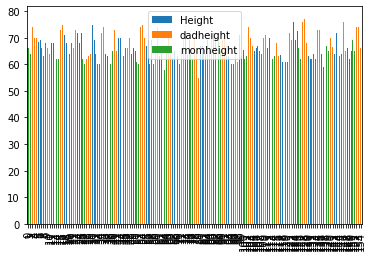
69.0 0.645161

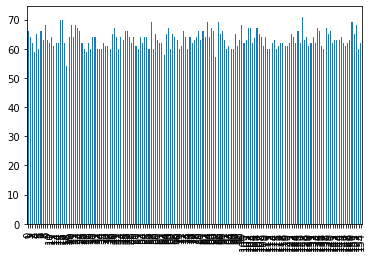
77.0 76.0 71.0 0.645161

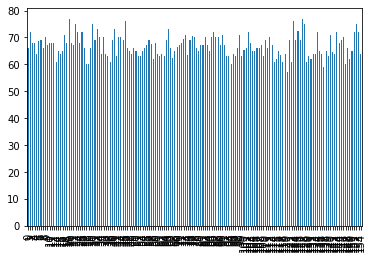
Length: 152, dtype: float64

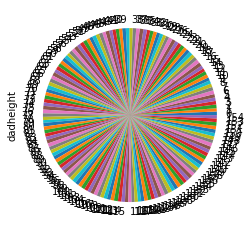
5.

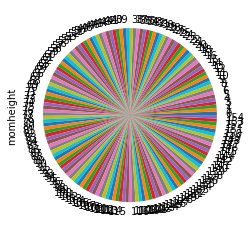


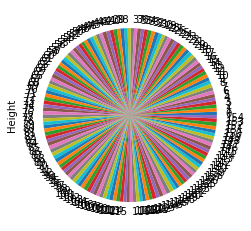












6. The frequency distribution of the height, dadheight and momheight are normally distributed

7.

computer 0.0 1.0 1.5 2.0 ... 37.0 40.0 50.0 84.0

TV 20.0 7.0 8.0 17.9375 ... 1.0 18.333333 10.0 1.0

[1 rows x 36 columns]

The frequency distribution between TV and computer is inversly proportional.

I.e the higher the computer frequency distribution, the lower the TV frequency distribution and vice versa.

8.

1.0 12.903226

10.0 10.322581

2.0 9.677419

4.0 7.096774

8.0 7.096774

15.0 7.096774

5.0 5.161290

20.0 4.516129

3.0 3.870968

14.0 3.870968

6.0 3.225806

0.0 3.225806

3.5 3.225806

1.5 1.935484

11.0 1.935484

9.0 1.935484

30.0 1.935484

35.0 1.290323

12.0 1.290323

21.0 1.290323

0.5 1.290323

40.0 0.645161

16.0 0.645161

25.0 0.645161

18.0 0.645161

13.0 0.645161

7.0 0.645161

5.5 0.645161

2.5 0.645161

100.0 0.645161

Name: TV, dtype: float64

10.0 12.987013

20.0 8.441558

5.0 8.441558

7.0 5.194805

2.0 5.194805

14.0 5.194805

25.0 4.545455

6.0 4.545455

15.0 4.545455

30.0 3.896104

3.0 3.896104

21.0 3.246753

35.0 2.597403

8.0 2.597403

4.0 2.597403

12.0 2.597403

1.0 2.597403

28.0 1.948052

40.0 1.948052

9.0 1.948052

17.5 1.298701

8.5 0.649351

12.5 0.649351

37.0 0.649351

0.0 0.649351

24.0 0.649351

7.5 0.649351

22.5 0.649351

17.0 0.649351

11.0 0.649351

32.0 0.649351

13.5 0.649351

50.0 0.649351

84.0 0.649351

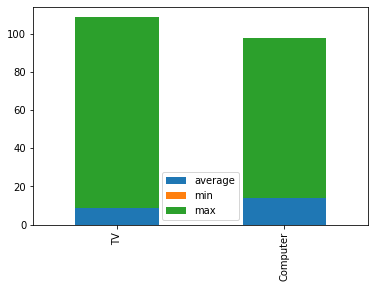
5.5 0.649351

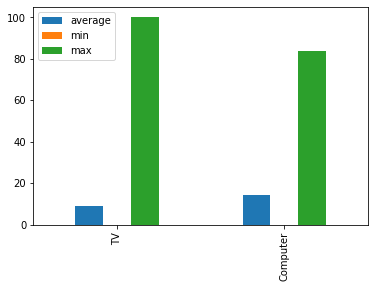
1.5 0.649351

Name: computer, dtype: float64

The distribution is inversly proportional.

9.



10. The stacked bar chart is a better visualization as it gives clear differences between the data compared. The other two comparisons may be hard to read for very large datasets.

11.

TV 8.741935

computer 14.068182

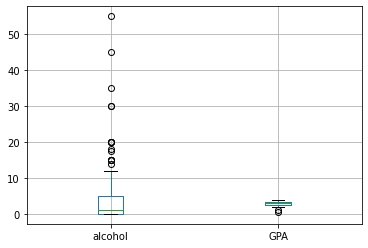
Sleep 6.985484

alcohol 4.409091

exercise 4.597403

GPA 2.913582

dtype: float64



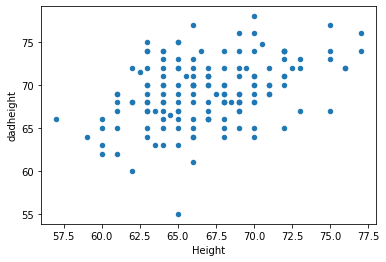
12.

13. TV:6.0, Computer: 10.0,Sleep:7.0,Alcohol:1.0,Height:66.5,MomHeight:63.0,DadHeight:69.0,Exercise:3.0,GPA:3.0

"""  
The medium of the variables compares as follows:  
 TV:6.0, Computer: 10.0,Sleep:7.0,Alcohol:1.0,Height:66.5,MomHeight:63.0,DadHeight:69.0,Exercise:3.0,GPA:3.0  
  
The IQR are as follows:  
 Sex:56.0  
 Seat:61.7  
 libarts:59.0  
 TV:53.0  
 computer:59.0  
 Sleep:58.0  
 alcohol:62.22  
 Height:58.5  
 momheight:65.0  
 dadheight:58.5  
 exercise:55.0  
 GPA:61.0  
   
The ranges are as follows:  
 Sex:67.5  
 Seat:72.0  
 libarts:74.0  
 TV:70.0  
 computer:67.5  
 Sleep:68.5  
 alcohol:74.5  
 Height:66.8  
 momheight:70.0  
 dadheight:67.0  
 exercise:65.0  
 GPA:69.0  
  
The upper quartile:  
 Sex:66.0  
 Seat:64.0  
 libarts:62.0  
 TV:59.0  
 computer:64.0  
 Sleep:60.0  
 alcohol:66.0  
 Height:63.0  
 momheight:68.0  
 dadheight:63.0  
 exercise:62.0  
 GPA:64.0  
   
The lower quartile:  
 Sex:10.0  
 Seat:2.3  
 libarts:3.0  
 TV:6.0  
 computer:5.0  
 Sleep:2.0  
 alcohol:3.78  
 Height:4.5  
 momheight:3.0  
 dadheight:4.5  
 exercise:7.0  
 GPA:3.0  
 """

14. Scatter plot

The height of the dad explains the height of the student.



15. Large number of students vs dad height comparison lies in the center to mean normal relation.

16. 90%: (6.759444916820999, 7.211522825114485)

95%:(6.716141844436291, 7.254825897499193)

99%: (6.631508445863302, 7.339459296072182)

17.

"""

The confidence interval increases with increase in levels ie.

90%: 0.4520779082934858

95%: 0.5386840530629016

99%: 0.7079508502088796

"""

18.

Sex Seat libarts TV computer Sleep alcohol Height momheight dadheight exercise GPA avgMomDadHeight

0 Female Back NonLib 13.0 10.0 3.5 12.0 66.0 66.0 71.0 10.0 4.00 68.5

1 Male Back NonLib 20.0 7.0 9.0 0.0 72.0 64.0 65.0 2.0 2.30 64.5

2 Male Back NonLib 15.0 15.0 6.0 0.0 68.0 62.0 74.0 3.0 2.60 68.0

3 Male Middle NonLib 8.0 20.0 6.0 0.0 68.0 59.0 70.0 6.0 2.80 64.5

4 Female Front NonLib 2.5 10.0 5.0 5.0 64.0 65.0 70.0 6.5 2.62 67.5

.. ... ... ... ... ... ... ... ... ... ... ... ... ...

150 Female Back NonLib 5.0 35.0 5.0 4.0 65.0 69.0 72.0 5.0 2.18 70.5

151 Male Back NonLib 8.0 32.0 6.0 3.0 72.0 65.0 71.0 1.5 2.80 68.0

152 Male Middle NonLib 9.0 7.0 7.0 10.0 75.0 68.0 74.0 7.0 3.50 71.0

153 Male Front NonLib 1.0 10.0 5.0 5.0 72.0 60.0 74.0 3.0 4.00 67.0

154 Female Middle NonLib 5.0 2.0 6.0 0.0 64.0 62.0 66.0 6.0 4.00 64.0

[155 rows x 13 columns]

19

"""

I chose the mean of dad + mom height to help relate to the Height column.

I get the average data by adding the momheight column to the dadheight column and dividing the sum by two.

"""

20.

avgHeights = pd.DataFrame([df\_new.Height, df\_new.avgMomDadHeight])

avgMeanHeight = avgHeights.mean()

"""

the overal mean range is small ie 0-4

The dad and mom height determines the Student height

"""

21.

ucdavis\_sample1 = df.sample(30)

22. DONE.