

## Group contract

**Group number:** CC06E1

**Name:** Alicia Molesworth

**GitHub link:** <https://github.sydney.edu.au/sshi4795/CC06E1>

### I agree to:

- Abide by the terms of this contract in relation to the group assessment for DATA2002/2902.
- Store all my written and code contributions to the assessment in the GitHub repository.
- Keep a record of my other contributions to the assessment (e.g. discussions, emails, meetings attended). A copy of this may be requested by the coordinator.
- Work together with respect for all members, acting honestly and ethically consistently and not demonstrating acts of bullying/harassment towards each other. (details in Student Charter.)
- Communication via two methods: FB messenger for quick communication and Zoom for group meetings/working on tasks.
- Try to check FB messenger everyday and check Github at least once a week (more as the deadline gets closer).
- A zoom call can be held at least once a week for meetings (possibly to assign tasks for the week), if more contact needed FB messenger can be used to organize another time in the week.

### I understand that:

- My agreement to these terms is indicated through the act of submitting this in Canvas.
- If I fail to meet my obligations as detailed in this group contract then I have failed to meet the assessment requirements for DATA2002/2902 and may be awarded a mark of zero for some or all of the project components.

## Exploratory data analysis

**Data set:** Body fat percentage

**Dependent variable:** Percentage body fat

Figure 1: Graph showing the missing observations in the dataset. All values are present.

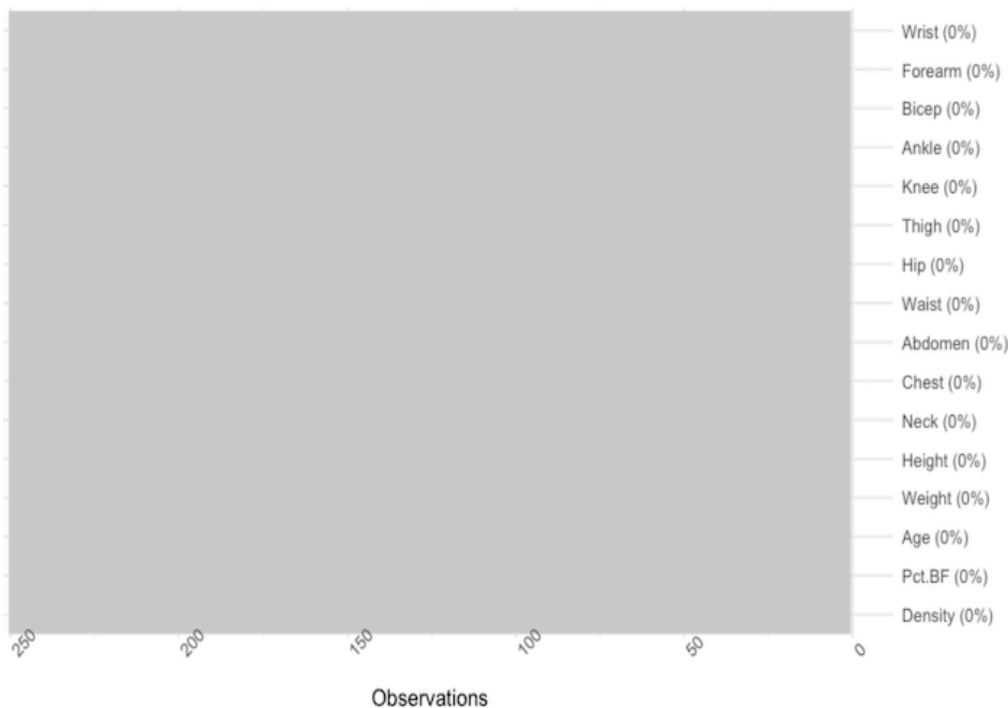


Figure 2: Descriptive statistics on variables grouped by age group.

	>70 (N=9)	20-29 (N=36)	30-39 (N=39)	40-49 (N=92)	50-59 (N=47)	60-69 (N=27)
<b>Pct.BF</b>						
Mean (SD)	24.7 (7.57)	14.2 (7.53)	17.6 (7.53)	19.6 (8.08)	18.7 (8.91)	24.2 (6.12)
Median [Min, Max]	27.0 [11.0, 33.6]	12.4 [3.70, 31.2]	17.7 [0.700, 34.3]	20.4 [0, 40.1]	16.1 [5.20, 47.5]	25.8 [11.8, 35.0]
<b>Weight_kg</b>						
Mean (SD)	79.5 (10.6)	81.2 (11.4)	81.2 (15.0)	81.2 (11.9)	78.8 (11.3)	82.0 (13.0)
Median [Min, Max]	77.4 [60.9, 94.1]	81.6 [60.4, 110]	80.4 [56.8, 109]	80.0 [53.7, 119]	76.9 [56.7, 104]	80.6 [57.8, 106]
<b>Height_m</b>						
Mean (SD)	1.75 (0.0425)	1.81 (0.0735)	1.79 (0.0586)	1.80 (0.0610)	1.77 (0.0714)	1.75 (0.0601)
Median [Min, Max]	1.77 [1.68, 1.79]	1.81 [1.64, 1.97]	1.80 [1.66, 1.89]	1.79 [1.65, 1.93]	1.77 [1.63, 1.97]	1.74 [1.67, 1.86]
<b>Abdomen</b>						
Mean (SD)	98.7 (9.17)	88.1 (9.16)	91.3 (10.6)	92.4 (10.1)	92.2 (9.68)	97.1 (10.3)
Median [Min, Max]	99.8 [83.6, 112]	87.2 [72.8, 107]	91.6 [74.6, 116]	90.4 [69.4, 126]	91.0 [76.0, 122]	97.8 [79.7, 118]
<b>Waist</b>						
Mean (SD)	38.8 (3.61)	34.7 (3.61)	35.9 (4.16)	36.4 (3.99)	36.3 (3.81)	38.2 (4.05)
Median [Min, Max]	39.3 [32.9, 43.9]	34.3 [28.7, 42.0]	36.1 [29.4, 45.6]	35.6 [27.3, 49.7]	35.8 [29.9, 48.1]	38.5 [31.4, 46.5]
<b>Hip</b>						
Mean (SD)	98.6 (5.43)	100 (6.13)	100 (7.80)	99.9 (6.48)	98.2 (5.51)	100 (6.54)
Median [Min, Max]	97.8 [88.8, 107]	100 [88.5, 114]	99.8 [85.3, 114]	99.5 [85.0, 126]	97.0 [87.5, 113]	99.6 [87.6, 114]
<b>BMI</b>						
Mean (SD)	26.0 (3.20)	24.8 (3.06)	25.3 (3.78)	25.0 (3.19)	25.2 (3.37)	26.8 (3.45)
Median [Min, Max]	24.6 [21.0, 30.1]	24.4 [19.2, 31.8]	25.1 [19.0, 33.2]	24.6 [18.0, 39.1]	25.1 [20.0, 37.6]	26.0 [20.7, 33.9]

Figure 3: Initial visualisations on the correlation between variables and their respective r correlation values.

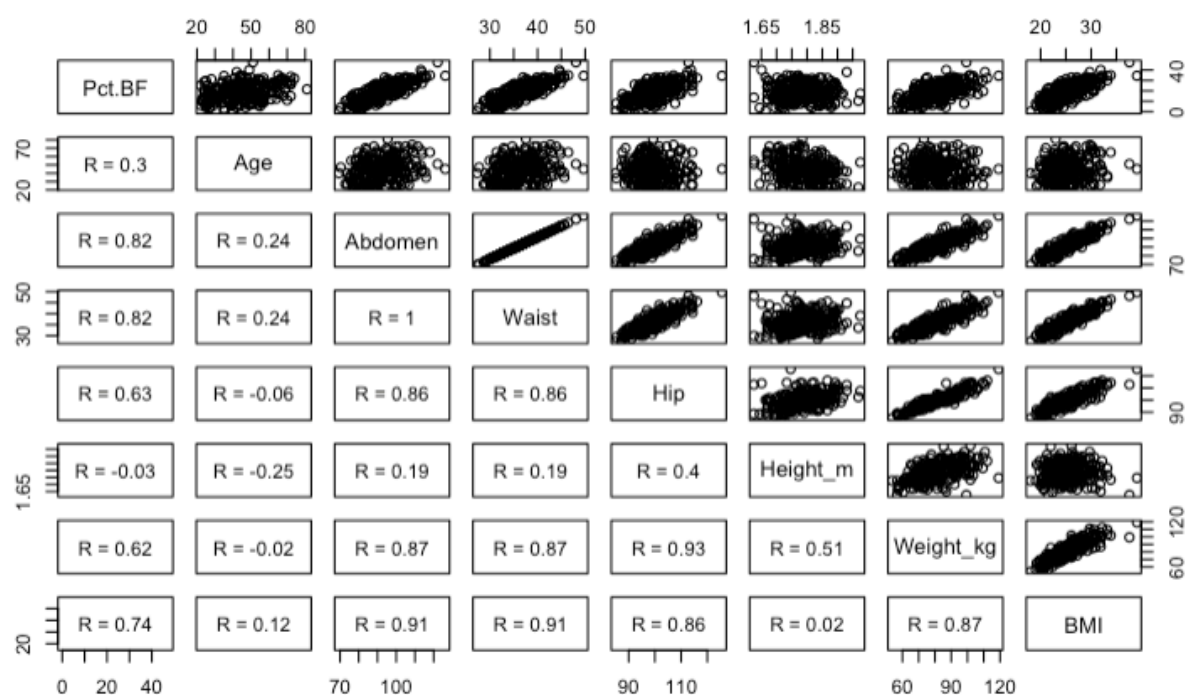
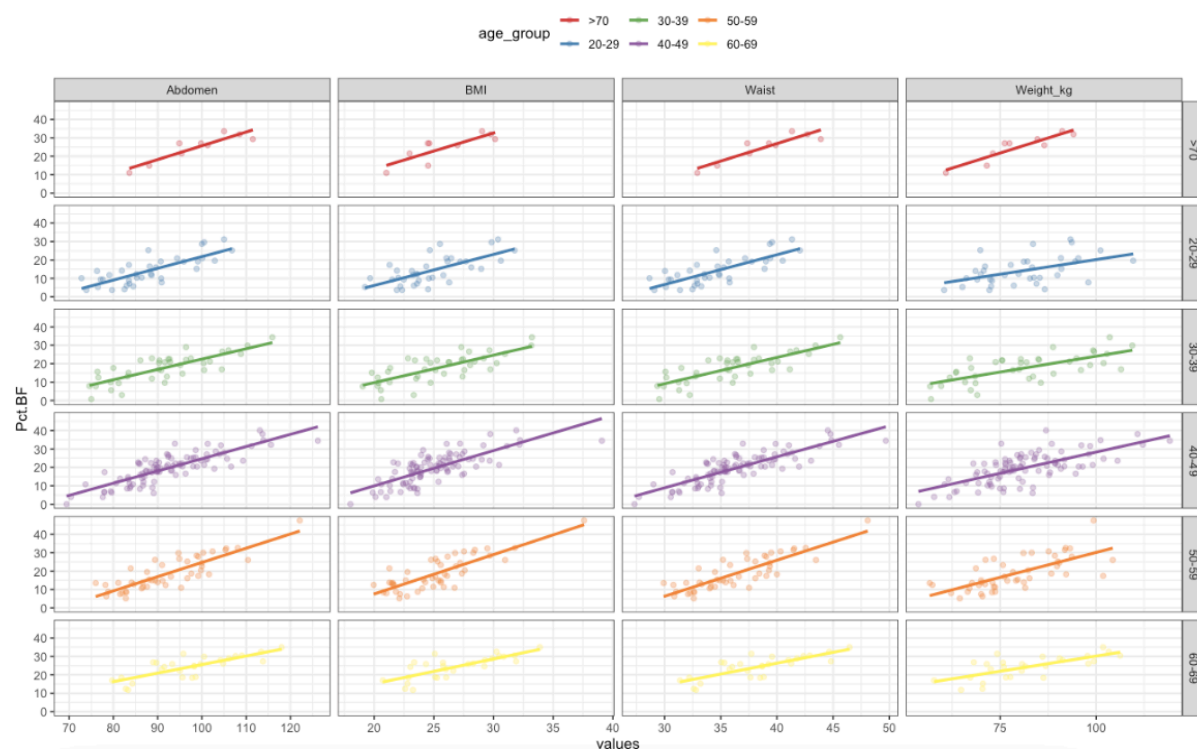


Figure 4: (Further visualisations) Combined scatterplots comparing multiple variables.



Summary: My initial EDA also included converting weight from pounds to kilograms, height from inches to metres, grouping the ages into decade categories and calculating BMI. From my initial EDA it looks like weight, waist and abdomen are important factors for predicting percentage body fat. There also appears to be a relationship between percentage body fat and BMI.