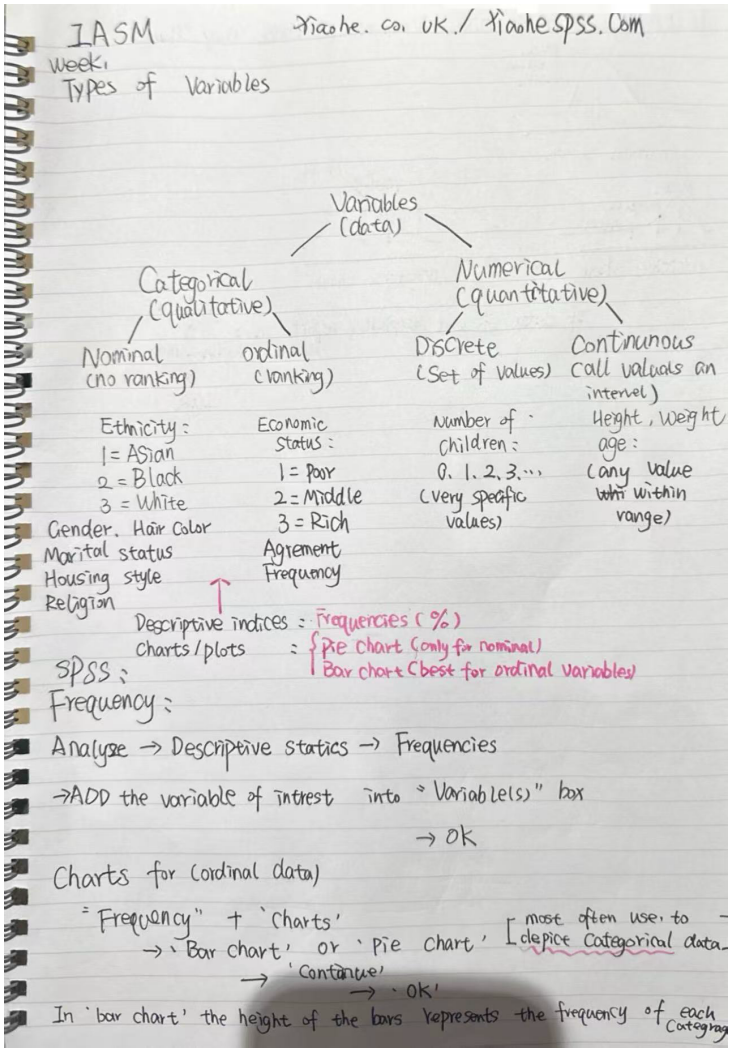
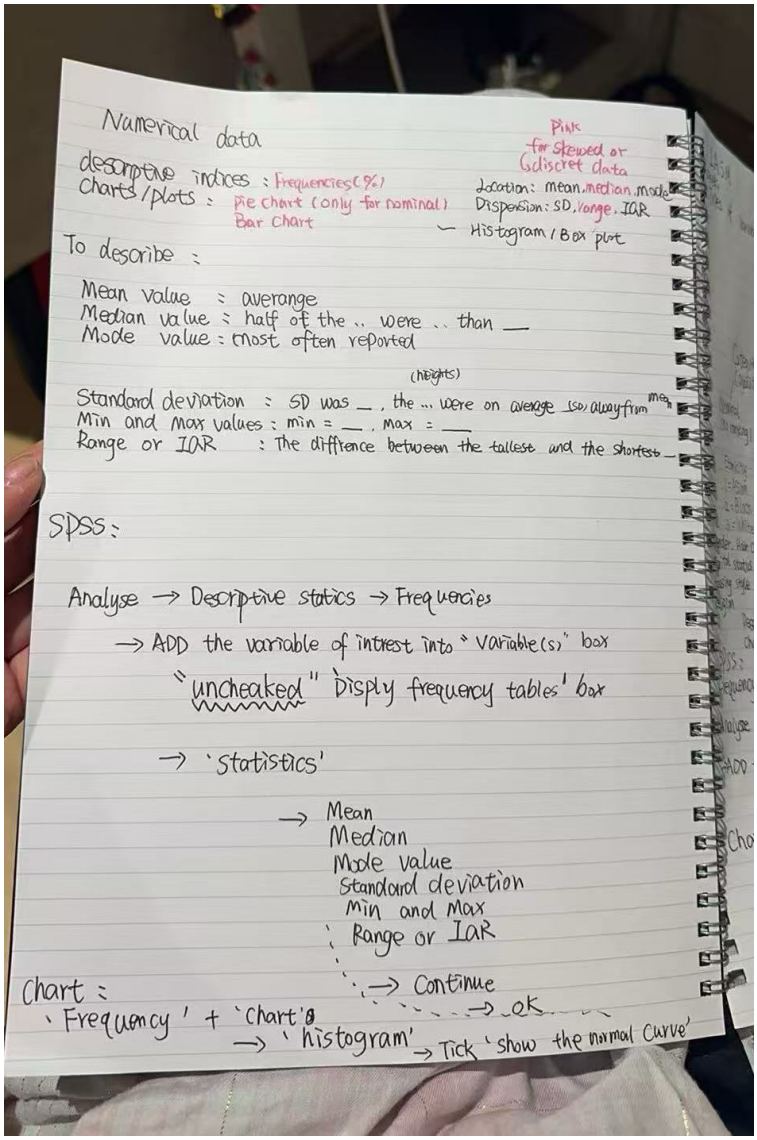
Notes:


IASM
xiaohe.co.uk / xiaohespss.com
week1
Types of Variables
Variables (data)
Categorical (qualitative)
Numerical (quantitative)
Nominal (no ranking)
ordinal (ranking)
Discrete (set of values)
Continuous (all values an interval)
Ethnicity:
1 = Asian
2 = Black
3 = White
Gender, Hair Color
Marital status
Housing style
Religion
Economic Status:
1 = Poor
2 = Middle
3 = Rich
Agrement
Frequency
Number of children:
0, 1, 2, 3, ...
(very specific values)
Height, Weight
age:
(any value within range)
Descriptive indices: Frequencies (%)
Charts/plots: Pie chart (only for nominal)
Bar chart (best for ordinal variables)
SPSS:
Frequency:
Analyse → Descriptive statics → Frequencies
→ADD the variable of intrest into "Variable(s)" box
→ OK
Charts for Cordinal data)
"Frequency" + 'Charts'
→ 'Bar chart' or 'Pie chart' [most often use, to depict Categorical data
→ 'Continue'
→ 'OK'
In 'bar chart' the height of the bars represents the frequency of each Categrag



Numerical data

descriptive indices : Frequencies(%)
charts/plots : pie chart (only for nominal)
Bar chart

Pink
for skewed or
Cdiscret data
location: mean, median, mode
Dispersion: SD, range, IQR
— Histogram / Box plot

To describe :

Mean value : average
Median value : half of the .. were .. than __
Mode value : most often reported

(heights)

Standard deviation : SD was __, the ... were on average 1SD away from mean
Min and Max values : min = __, Max = __
Range or IQR : The diffrence between the tallest and the shortest

SPSS:

Analyse → Descriptive statics → Frequencies

→ ADD the variable of intrest into "variable(s)" box

"uncheaked" 'Disply frequency tables' box

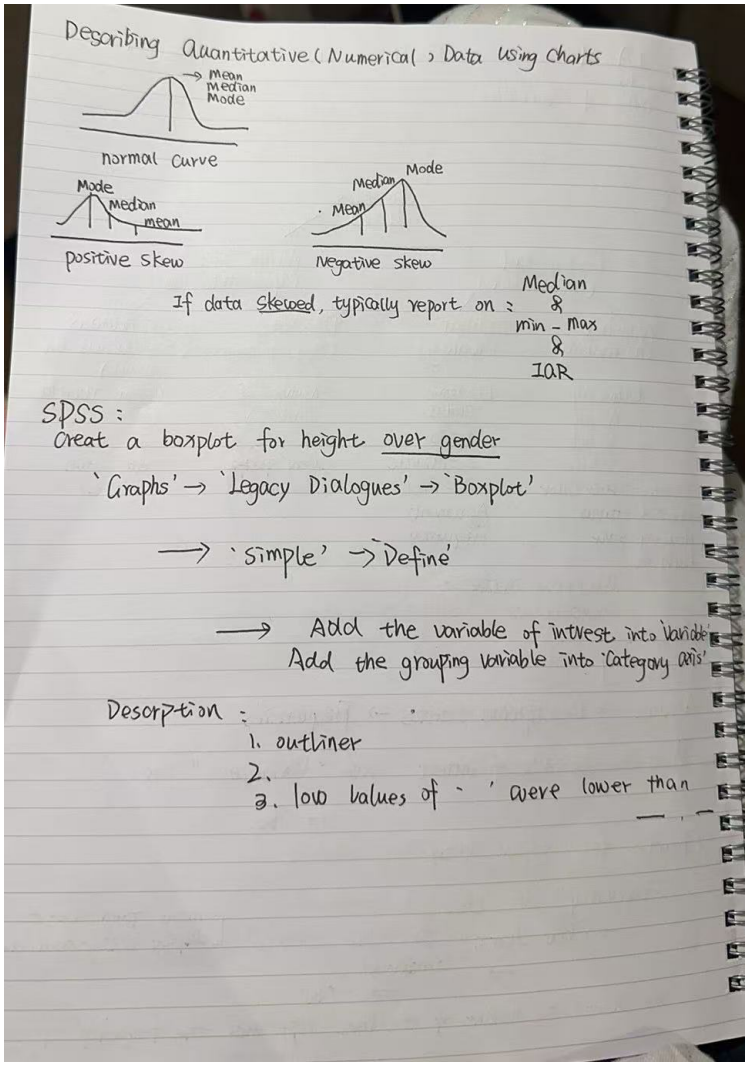
→ 'statistics'

→ Mean
Median
Mode value
Standard deviation
Min and Max
Range or IQR

→ Continue

→ ok

Chart :
'Frequency' + 'chart's
→ 'histogram' → Tick 'show the normal curve'



✅ Question 1

Q: The variable is categorical. What chart should you use?
A: Use a **Bar Chart**.
📌 *Why?* Bar charts are best for showing the frequency of categories (like gender, blood type).

✅ Question 2

Q: What does 54.2% refer to in SPSS output for gender?
A: That's the **Valid Percent** — the percentage *excluding* missing responses.
📌 *Why?* It shows the percentage **only among people who answered** the question.

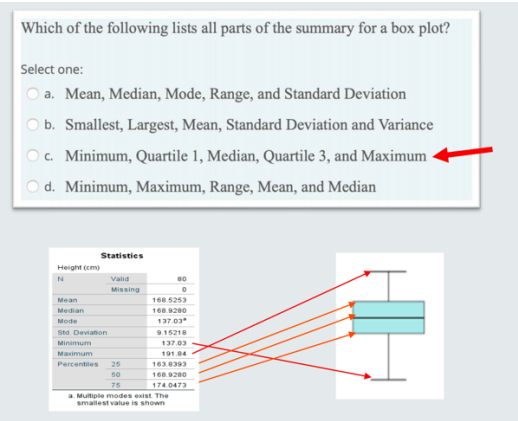
✅ Question 3

Q: What happens in a normal distribution?
A: The **mean, median, and mode are equal**.
📌 *Why?* Normal distribution is symmetrical — most values cluster around the center.

✅ Question 4

Q: What's the best way to describe a skewed variable?
A: Use the **median and min–max**, not the mean and SD.
📌 *Why?* Mean and SD are sensitive to outliers and skewness — median is more stable.

✅ Question 5



✅ Question 6

Q: Interpreting the "pleasant" and "unpleasant" boxplots:
A:

- The **"pleasant"** group has an **outlier**.
- Both distributions are **skewed**.
- The **unpleasant** group has a **higher median**.
- The **pleasant** group has **more variability**.
  📌 *Why?* Boxplots visually show skew, spread, medians, and outliers.

📊 Categorical Variables (Qualitative)

Variable Name	Type	Notes
Gender	Nominal	Male / Female / Other — no order
Ethnicity	Nominal	Race categories — no order
Marital Status	Nominal	Single, Married, etc. — no order
Employment Status	Nominal	Employed, Unemployed, etc.
Diagnosis	Nominal	e.g. ADHD, Anxiety, None
Satisfaction Level	Ordinal	e.g. Very satisfied → Very dissatisfied
Education Level	Ordinal	Primary, Secondary, University
Pain Level	Ordinal	Mild, Moderate, Severe
Social Class	Ordinal	I, II, III, etc. — has a rank
Likert Scale Items	Ordinal	Agree → Disagree type responses

🔢 Numerical Variables (Quantitative)

🔷 Interval Variables

Variable Name	Type	Notes
IQ Score	Interval	No true zero
Temperature (°C/°F)	Interval	0 doesn’t mean “no temperature”
Calendar Year	Interval	0 AD isn’t “no time”

🔶 Ratio Variables (True Zero Exists)

Variable Name	Type	Notes
Age	Ratio	0 = birth, can say “twice as old”
Height / Weight	Ratio	0 = none, meaningful ratios
Income	Ratio	£0 = no money
Hours of Sleep	Ratio	Can be 0 hours
Heart Rate	Ratio	Beats per minute
LDL Cholesterol	Ratio	Measurable biological marker
Number of Children	Ratio	Countable whole numbers
Years Lived in London	Ratio	0 = just moved

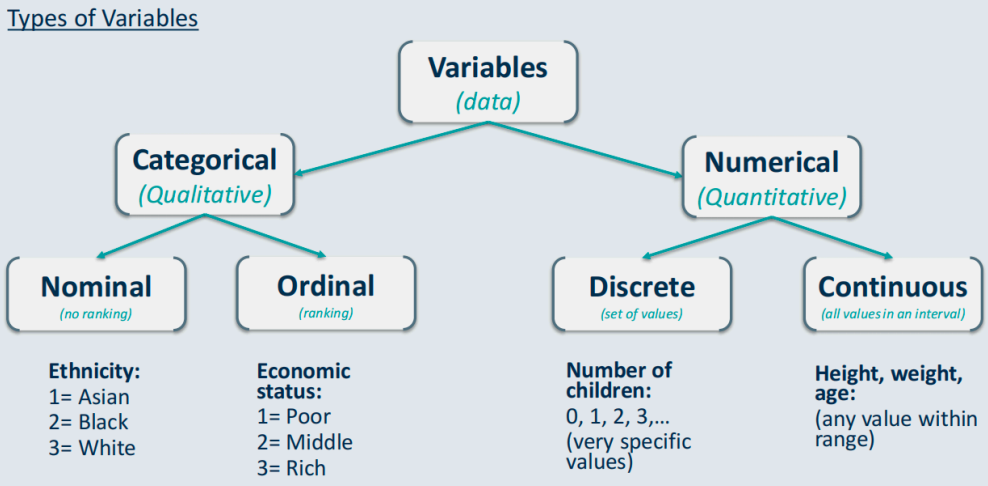
Tip for SPSS:

If a variable has **value labels**, it’s often **categorical** (even if stored as numbers). If it’s **measured in real units or amounts**, it’s likely **numerical** (interval or ratio).

An icon next to each variable provides information about data type and level of measurement.

	Numeric	**String**	**Date**	**Time**
Scale (Continuous)		n/a		
Ordinal				
Nominal				

Types of Variables



Categorical DATA

Categorical Data: nominal or ordinal?

Nominal data can't be expressed as a number and can't be measured. They are **names** which represent qualities of the observations, characteristics, categories the observations belong to.

Nominal data can take on numerical values (example: 1 for male, 2 for female, 3 for other) but those numbers don't have mathematical meaning - are coded for ease of computation in most statistical software. Ordering has no meaning.

Ethnicity
i. Asian
ii. Black
iii. White
iv. Other

Gender
a) Cis man
b) Cis woman
c) Trans man
d) Trans woman
e) Other

Hair colour
1. Blonde
2. Brown
3. Brunette
4. Red

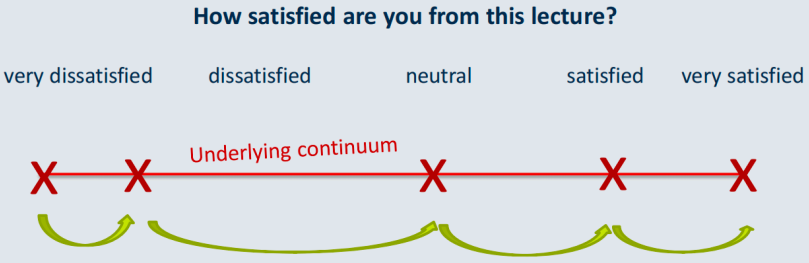
Marital Status
- Married
- Single
- Widowed
- Self-partnered

Housing Style
- ❑ Detached
- ❑ Semi-Detached
- ❑ Terraced
- ❑ Bungalow
- ❑ Flat

Religion
I. Buddhism
II. Christianity
III. Hinduism
IV. Islam
V. No religion

Categorical Data nominal or ordinal?

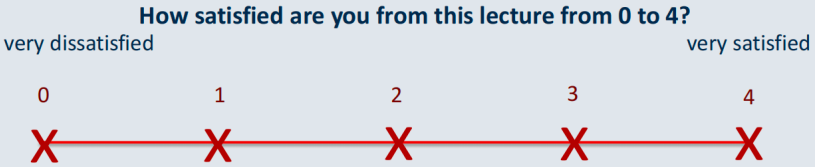
Ordinal data take on numerical values and those numbers represent the **order** of the categories. However they lack mathematical meaning as the spacing between categories is not necessarily equal.



Ordinal (categorical) data or Interval (numeric data?)

Ordinal data take on numerical values and those numbers represent the **order** of the categories. However they lack mathematical meaning as the spacing between categories is not necessarily equal.

But if the variable is structured in a way that it is clear that the spacing is equidistant, and differences between then are meaningful, then the data are **interval** data (numerical data). An example:



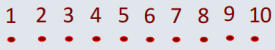
That is because it now makes sense to say 4, is double as 2 and the distance between 1 and 3 is the same as, say, 2 and 4. There is a mathematical underpinning in the numbers now.

Numerical Data

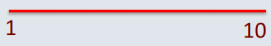
Numerical Data

Sometimes it can also be tricky to tell apart **discrete** and **continuous** data. Discrete data take only very specific (and pre-specified) **set** of values. Continuous data can take all values in a prespecified **interval**.

Discrete {1, 2, 3, 4, 5, 6, 7, 8, 9,10}



Continuous [1,10]



Typically, discrete data are **counts** and continuous data are **measurements**.

How many children?	Weight
How many cars?	Height
How many times?	Age

Knowledge check:

Knowledge Check Solutions

1. **Which of the variable(s) are classified as quantitative variable(s)?**
   Age, Height, LDL, Number of Children
   These variables take numerical values only and the values reflect the actual measurement (with units) of the subjects or objects we are measuring.
2. **Which of the variable(s) are classified as qualitative variable(s)?**
   Blood Group, Gender, Feeling Happy, Smoke, Social class
   These variables are represented by categories and each category represents a particular characteristic of interest within a group of subjects or objects.
3. **Which of the variable(s) are classified as nominal variable(s)?**
   Gender, Blood Group, Smoke
   These variables consist of categories that are mutually exclusive but have no ranked order, e.g. Male / Female.
4. **Which of the variable(s) are classified as ordinal variable(s)?**
   Feeling Happy, Social Class
   These variables consist of categories that are mutually exclusive and have a ranked order. Thus, for example, the category "strongly agree" may precede "agree". Note that the "interval" between categories may not be numerically equal.
5. **Which of the variable(s) are classified as discrete variable(s)?**
   ID, Number of Children
   These variables take integer values. ID is the subject or case number and Number of Children are counts.
6. **Which of the variable(s) are classified as continuous variable(s)?**
   Age, LDL, Height
   These variables can take any value within an interval, including decimal parts. The precision of the measurement will depend on the measuring device used.

Knowledge Check Solutions

Q1. Which of the variables would you describe using **frequencies (Percentages %)**
Blood Group, Gender, Feeling Happy, Smoke, Social class.
All of these variables are qualitative (categorical) variables and would be described by frequencies and percentages.

Q2. Which of the variable(s) would you use a **pie chart?**
You could use a pie chart or bar chart to visualise any of the above variables, but it may be more meaningful, visually, to do a pie chart for where we have more than 2 categories like blood group. For the ordinal variables is best to use the bar charts (feeling happy, social class)

Q3. Below is a frequency distribution for the variable social class give an interpretation of this information.
In our sample, half of the individuals were in social classes III to IV (N=6, 50%).

Q4. Below is a bar chart of the variable 'Blood Group' what does the chart show us?
The majority of subjects belong to blood groups A and B (N_A = 3, N_B = 3, 60%) with the rest of the subjects split evenly between blood groups O and AB (N_O = 2, N_{AB} = 2, 40%)

🧠 Why Do We Use Statistics?

- We can't predict what **one person** will do, but we can understand how **groups of people** behave.
- Statistics helps us explain **why people differ** — in age, health, mood, etc.
- It lets us **make guesses about the population** by studying a **sample**

🔍 What Is a Variable?

- A **variable** is something that can change from person to person.
- Examples: age, gender, mood, number of children, blood pressure.
- Variables help us describe and compare people.

🧱 Types of Variables

There are **two main types**:

1. **Categorical (Qualitative)**
    - **Nominal:** Just names or labels (no order). E.g. blood group, gender.
    - **Ordinal:** Categories with a clear order. E.g. satisfaction levels (very satisfied → dissatisfied).
2. **Numerical (Quantitative)**
    - **Discrete:** Countable numbers. E.g. number of kids.
    - **Continuous:** Can take any value (even decimals). E.g. height, weight, age.

How to Describe Categorical Data

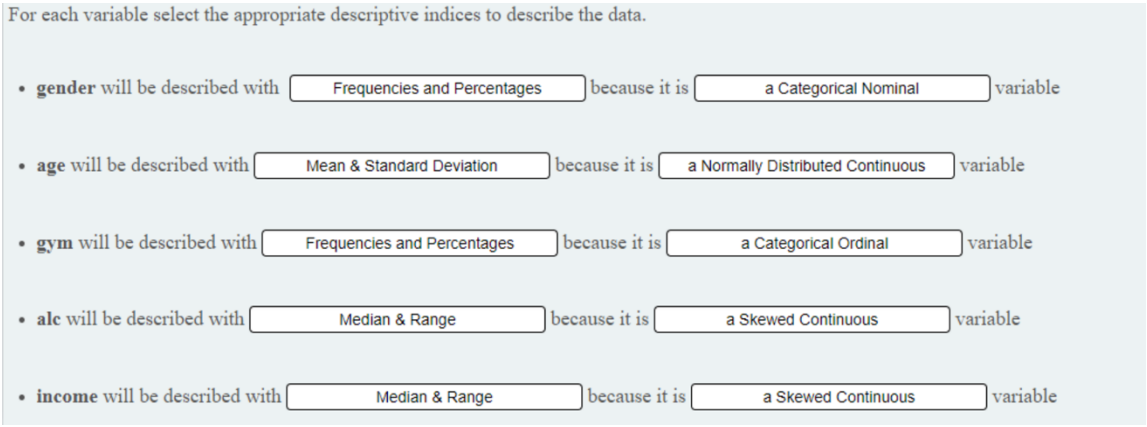
- We **count how many** people fall into each category — that's called **frequency**.
- We also show this in **percentages**.
- Charts:
  - **Bar chart** for nominal and ordinal data.
  - **Pie chart** is for nominal data only.

📈 How to Describe Numerical Data

- Instead of counts, we use **summary measures**:
  - **Mean** = average
  - **Median** = middle value
  - **Mode** = most common value
  - **Standard Deviation (SD)** = how spread out the data is
  - **Range** = max - min
  - **IQR (Interquartile Range)** = difference between Q3 and Q1
- Charts:
  - **Histogram** = for showing distribution of numerical data.
  - **Boxplot** = for comparing groups.

Skewed vs Normal Data

- **Normal Distribution**: Mean ≈ Median ≈ Mode. Use **mean and SD**.
- **Skewed Data**: Use **median, min-max, and IQR**.
- **Categorical: Use Frequencies and Percentages**



- **💻 SPSS Tips**

- To check variable summaries:
  `Analyze > Descriptive Statistics > Frequencies`
  (Uncheck frequency table if using numerical data)
- To create graphs:
  `Graphs > Legacy Dialogs > Bar / Pie / Histogram / Boxplot`
- Use SPSS outputs to spot typos, outliers, or incorrect values.

🔢 When to Use `Frequencies`

✅ **Best for Categorical Variables** (nominal or ordinal)

When to Use `Descriptives`

✅ **Best for Numerical Variables** (continuous or discrete)

⚖️ Quick Comparison:

Feature	Frequencies	Descriptives
Use for	Categorical or ordinal data	Continuous (scale) data
Output	Frequencies + % + charts (bar/pie)	Mean, SD, Min, Max
Charts included?	Yes (pie, bar)	No (you need to use Graphs separately)
Can spot typos?	Yes	No, but good for checking spread of data

Lecture slides How to: W1 L3-4

Qualitative (Categorical) Data

In categorical data, one would be interested in how many people are in each category and in total. We call this the '**frequency** of each category' and we use 'N' to symbolise the number of people. We also express these frequencies as **percentages** (%). Let's look at Gender (nominal data) as an example

Table 1: SPSS Frequency table for Gender

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	28	35.0	36.4	36.4
	Female	30	37.5	39.0	75.3
	Other	19	23.8	24.7	100.0
	Total	77	96.3	100.0	
Missing	System	3	3.8		
Total		80	100.0		

categories

missing values

Number of people in each category

Totals with and without missing values

% with and without missing values

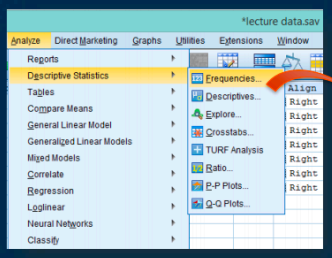
- 35% of the individuals in the sample (N=80) identified themselves as males
- 36.4% of the individuals who responded (N=77) identified themselves as males
- 75.3% of the individuals who responded (N=77) identified themselves as either males or females.

The cumulative % makes more sense in ordinal data

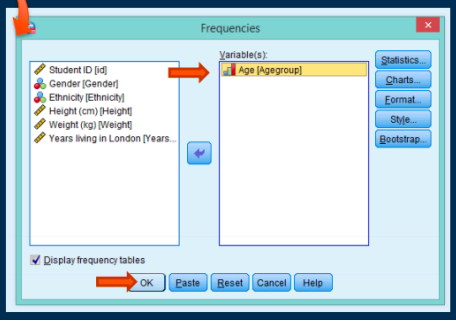
SPSS Slide: 'How to' Steps

You can create the **frequency table** for agegroup (ordinal data) using the following steps:

Click on the 'Analyse Tab'→ 'Descriptive Statistics' → 'Frequencies'



Add the variable of interest (agegroup) into the 'Variable(s)' box



Click on 'OK.

Output and Interpretation

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Up to 20 years old	19	23.8	23.8	23.8
	21 to 25 years old	45	56.3	56.3	80.0
	26 to 30 years old	12	15.0	15.0	95.0
	31 years old and above	4	5.0	5.0	100.0
	Total	80	100.0	100.0	

INTERPRETATION: In our sample, most people belong to the 21-25 years old **age** group (N=45, 56.3%). The vast majority of the individuals in our sample were up to 25 years old (N=64, 80.0%). Only 4 people (5.0%) were 31 years old or above.

Ethnicity

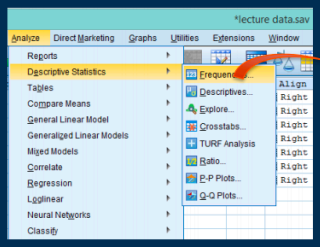
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Black	11	13.8	13.8	13.8
	White	19	23.8	23.8	37.5
	Asian	17	21.3	21.3	58.8
	Mixed	18	22.5	22.5	81.3
	Other	14	17.5	17.5	98.8
	22.00	1	1.3	1.3	100.0
	Total	80	100.0	100.0	

Typo spotted

By creating a frequency table for Ethnicity we were able to spot a typo/error in the data.

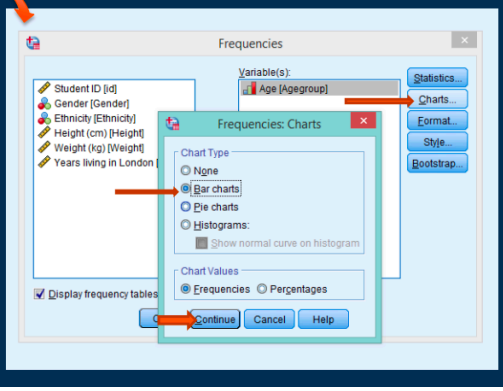
SPSS Slide: 'How to' Steps

You can create the **charts** for agegroup (ordinal data) using the following steps:



Click on the 'Analyse Tab'→ 'Descriptive Statistics' → 'Frequencies'

Add the variable of interest (agegroup) into the 'Variable(s)' box

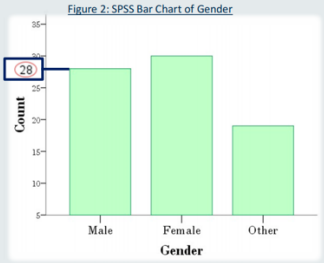


Click on 'Charts'

Choose 'Bar Chart' or 'Pie Chart'
Click 'Continue'
Click on 'OK'.

- To depict categorical data, most often we use a **Bar Chart** or a **Pie Chart:**

Figure 2: SPSS Bar Chart of Gender



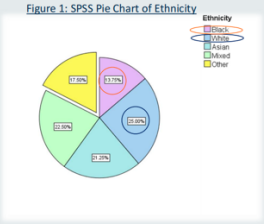
Gender

	Frequency	Percent
Male	28	35.0
Female	30	37.5
Other	19	23.8

In a bar chart, the height of the bars represents the frequency of each category.

- To depict categorical data, most often we use a **Bar Chart** or a **Pie Chart:**

Figure 1: SPSS Pie Chart of Ethnicity



Ethnicity

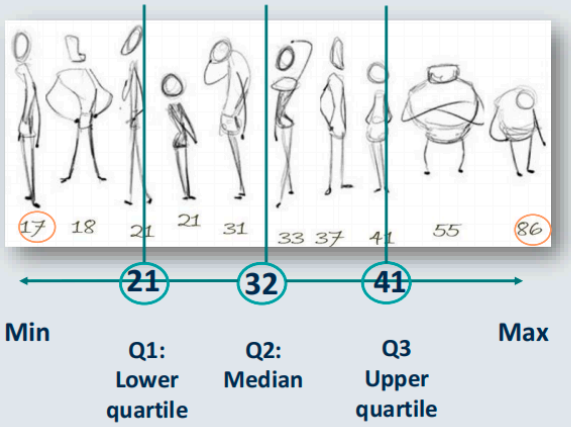
	Frequency	Percent
Black	11	13.8
White	20	25.0
Asian	17	21.3
Mixed	18	22.5
Other	14	17.5

only for nominal data

In a pie chart, the size of the sector represents the frequency of each category. More people, more pie.

Quantitative (Numerical) Data

Other measures that are useful to describe numerical data are called **Quartiles.**



Lower quartile = 21

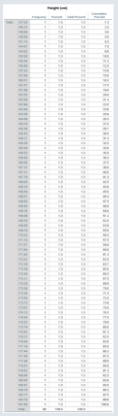
Median= 32

Upper quartile = 41

Interquartile Range = 41 – 21
= 20

To describe the metrical or numerical variable, we need to properly summarise it.

Instead of reporting this



We understand more by reporting on:

Measures of location (central tendency)

- Mean value: the **average** height of the students was 168.5cm (5.5ft)
- Median value: half of the students **were taller** than 169cm (5.5ft)
- Mode value: the height **most often** reported was 173cm (5.7ft)

Measures of dispersion (spread, variability)

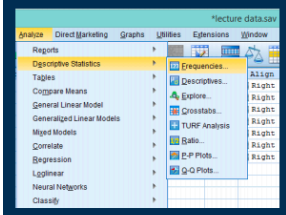
- Standard deviation: SD was 9cm (0.3ft): the heights were on average 9cm away from the mean height of 168.5cm
- Min and max values: **min** height =137cm (4.5ft), **max** height=192cm (6.3ft)
- Range or IQR The difference between the tallest and the shortest student was 10 cm (0.3ft)

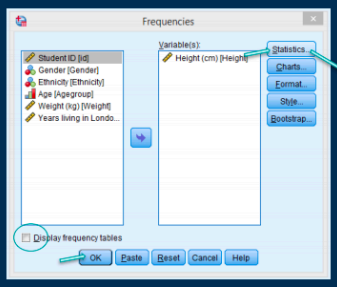
Descriptive indices:

SPSS Slide: ‘How to’ Steps

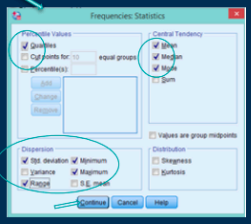
You can create the descriptive indices for height using the following steps:
Click on the ‘Analyse Tab’→ ‘Descriptive Statistics’ → ‘Frequencies’

Add the variable of interest (height) into the ‘Variable(s)’ box
Make sure the ‘Display frequency tables’ box is unchecked





Click on ‘Statistics’ and choose the indices you want to report.



Instead of frequencies we now want measures of central tendency (location) and measures of dispersion (spread).

Click on ‘Continue’
Click on ‘OK

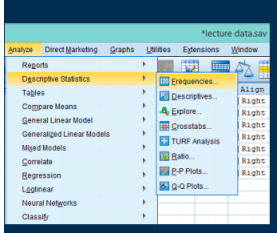
Chart:

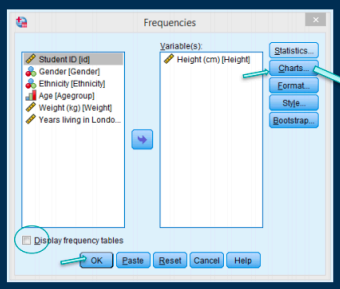
You can create a chart using the following steps:

Click on the 'Analyse Tab'→ 'Descriptive Statistics' → 'Frequencies'

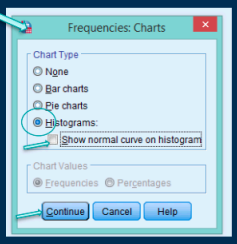
Add the variable of interest (height) into the 'Variable(s)' box

Make sure the 'Display frequency tables' box is unchecked





Click on 'charts' and choose the chart you want to report.



For the numerical variable height we would prefer the histogram

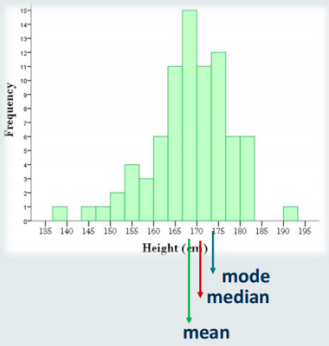
Tick 'show the normal curve'

Click on 'Continue'

Click on 'OK'

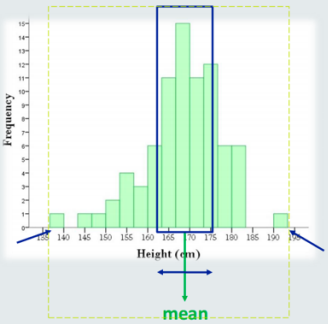
Measures of location (central tendency): they show where about most of the values are

Statistics		
Height (cm)		
N	Valid	80
	Missing	0
Mean		168.5750
Median		169.0000
Mode		173.00
Std. Deviation		9.16760
Range		55.00
Minimum		137.00
Maximum		192.00



Measures of dispersion (spread): they show how variable the values are

Each person obviously has a value within the interval [min, max]



Statistics		
Height (cm)		
N	Valid	80
	Missing	0
Mean		168.5750
Median		169.0000
Mode		173.00
Std. Deviation		9.16760
Range		55.00
Minimum		137.00
Maximum		192.00

People with values within the interval [mean-sd, mean+sd].

Box chart:

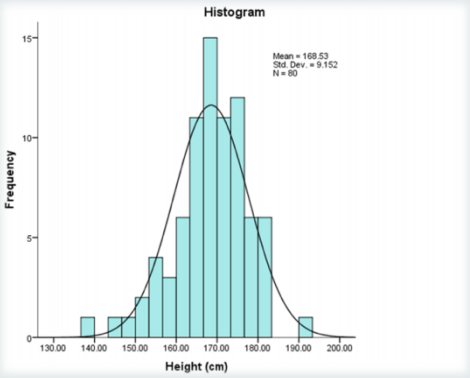
A chart has all the information we need and is easier to understand

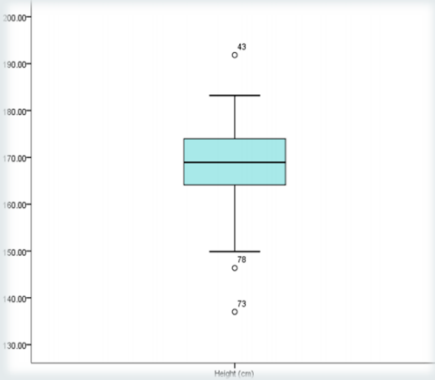
Statistics

Height (cm)

N	Valid	80
	Missing	0
Mean		168.5253
Median		168.9280
Mode		137.03[a]
Std. Deviation		9.15218
Minimum		137.03
Maximum		191.84
Percentiles	25	163.8393
	50	168.9280
	75	174.0473

a. Multiple modes exist. The smallest value is shown





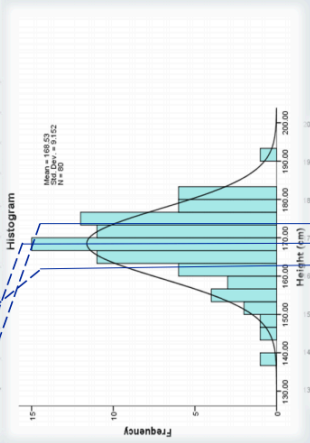
To describe a numerical variable, we need to properly **summarise** it.

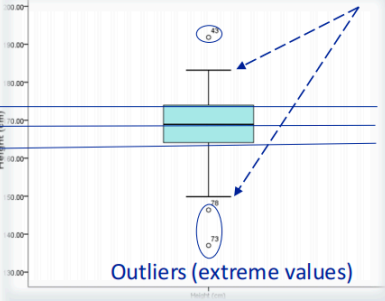
Statistics

Height (cm)

N	Valid		80
	Missing		0
Mean			168.5253
Median			168.9280
Mode			137.03[a]
Std. Deviation			9.15218
Minimum			137.03
Maximum			191.84
Percentiles	25	Q1	163.8393
Quartiles	50	Q2	168.9280
	75	Q3	174.0473

a. Multiple modes exist. The smallest value is shown





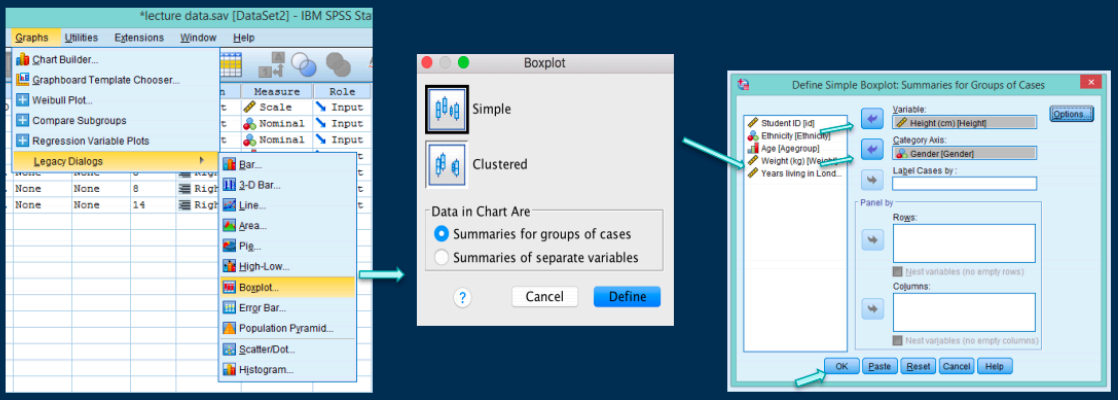
Min and max if outliers were omitted

Outliers (extreme values)

How to do the box chart:

You can create the boxplot for height **over gender,** using the following steps:

Click on 'Graphs'→ 'Legacy Dialogues'→ 'Boxplot'

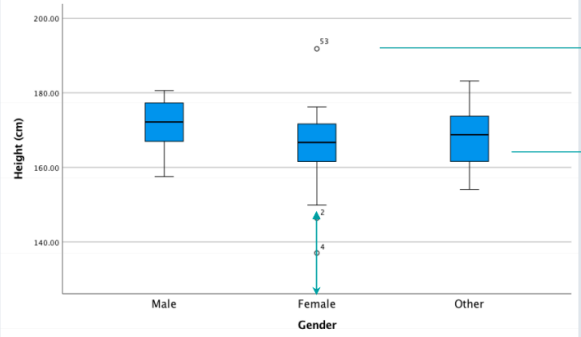


Choose a 'simple' layout and click 'Define'

Add the variable of interest (height) into the 'Variable(s)' box
Add the grouping variable (gender) into the 'Category axis' box

Click on 'OK'

The box plot is very useful in comparing groups visually.



Outliers were 'females'

The values of 'other' completely cover the values of 'males'.

The low values of 'females' were lower than all the values of 'males' and 'other'.

Practical Quiz:

Welcome to the Topic 1 Practical Quiz

For this quiz, the data set consists of n=120 members of a gym. The variables being measured are:

· **gender**: the reported gender (1: male, 2: female, 3: other)

· **age**: the member's age (in years)

· **gym**: how often they visit the gym (1=once per week, 2=twice per week, 3=three or more times per week)

· **alc**: alcohol units per week

· **income**: their gross annual income (1:up to £10,000, 2: £10,001 to £20,000, 3: £20,001 to £30,000, 4: £30,001 to £40,000, 5: £40,001 to £50,000, 6: more than £60,000)

- Age: Bar Chart ✗ *[Histogram]*
- Alc: Histogram ✓
- Gender: Bar Chart ✓
- Gym: Bar Chart ✓
- Income: Histogram ✓

- gender is a Categorical ✓ Nominal ✓ variable
- age is a Numerical ✓ Continuous ✓ variable
- gym is a Categorical ✓ Interval ✗ *[Ordinal]* variable
- alc is a Categorical ✗ *[Numerical]* Ordinal ✗ *[Continuous]* variable
- income is a Numerical ✓ Continuous ✗ *[Interval]* variable

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	48	40.0	40.7	40.7
	female	64	53.3	54.2	94.9
	other	6	5.0	5.1	100.0
	Total	118	98.3	100.0	
Missing	System	2	1.7		
Total		120	100.0		

Use "percent", to describe.

The age of the individuals in our sample ranged from 19.7 to 29.6 ✓ years old with mean age=24.7 ✓ (sd=1.95 ✓)years.

In our sample, there were 64 ✓ individuals who identified themselves as females (54.2 ✗ *[53.3]* %), 48 ✓ individuals who identified themselves as males (40.7 ✗ *[40.0]* %), and 6 ✓ individuals who chose "other" as their gender identity (5.1% ✗ *[5.0%]*). There were 2 ✓ individuals who did not respond to this question (1.7 ✓ %).

Most individuals earned between £30001 and £40000 ✓ a year. Half of the individuals earned more than ✓ £30001 a year. 10.0 ✗ *[10.2]* % of those who responded earned more than £60000.

Institute of Psychiatry, Psychology and Neuroscience

Module Title: Introduction to Statistics

Session Title: Practical Quiz 1

Topic title: Measurement and graphical representations of data

Practical Quiz 1

Welcome to the Topic 1 Practical Quiz

For this quiz, the data set consists of n=120 members of a gym. The variables being measured are:

- **gender**: the reported gender (1: male, 2: female, 3: other)
- **age**: the member's age (in years)
- **gym**: how often they visit the gym (1=once per week, 2=twice per week, 3=three or more times per week)
- **alc**: alcohol units per week
- **income**: their gross annual income (1:up to £10,000, 2: £10,001 to £20,000, 3: £20,001 to £30,000, 4: £30,001 to £40,000, 5: £40,001 to £50,000, 6: more than £60,000)

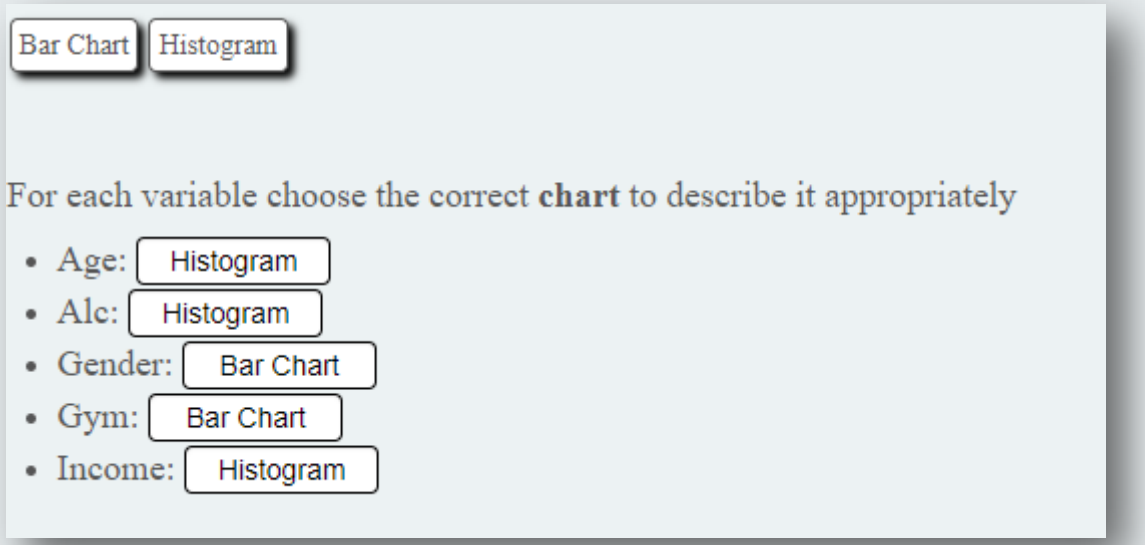
Question 1

Ordinal | Continuous | Interval | Categorical | Numerical | Nominal

To begin with, identify the type of variables in the dataset. Please indicate whether the variable is **'Categorical'** or **'Numerical'** first. Then the type of variable: **'Nominal, Ordinal, Interval, Discrete, Continuous'**

- gender is a Categorical / Nominal variable
- age is a Numerical / Continuous variable
- gym is a Categorical / Ordinal variable
- alc is a Numerical / Continuous variable
- income is a Numerical / Interval variable

Question 2


Bar Chart
Histogram
For each variable choose the correct **chart** to describe it appropriately
Age: Histogram
Alc: Histogram
Gender: Bar Chart
Gym: Bar Chart
Income: Histogram

Question 3

range=9.90 | normally | symmetrical | 24.74 | 24.50 | 2 | 22.70 | non-symmetrical | SD=1.95

Check the data for typos, note them down and remove them. Then, describe the variable **age** and comment on its distribution, giving answers to 2 decimal places.

There was one typo for the variable 'age', of value [2].

The average age was [24.74] ([SD=1.95]) years. The variable 'age' is [normally] distributed. We can see in the histogram that it is [symmetrical] around the mean.

Question 4

non-normally | normally | below | symmetrical | non-symmetrical | above

Please comment on the distribution of **alc**.

The variable 'alc' is [non-normally] distributed. We can see in the histogram that it is [non-symmetrical] around the mean. Most of the values are [below] 10 units of alcohol.

Question 5

0 100% 3 27.5%

Please check the variable **'gym'** for typos, note them down and remove them if there were any. Comment on what you observe in the values of **'gym'**.

There were 3 typos in the database for this variable.

Among the people in the sample, 27.5% responded that they visit the gym three times or more in a week.

Question 6

a Categorical Nominal | a Normally Distributed Continuous | a Categorical Ordinal | Mean & Standard Deviation | Median & Range | Frequencies and Percentages | a Skewed Continuous

For each variable select the appropriate descriptive indices to describe the data.

- **gender** will be described with [Frequencies and Percentages] because it is [a Categorical Nominal] variable
- **age** will be described with [Mean & Standard Deviation] because it is [a Normally Distributed Continuous] variable
- **gym** will be described with [Frequencies and Percentages] because it is [a Categorical Ordinal] variable
- **alc** will be described with [Median & Range] because it is [a Skewed Continuous] variable
- **income** will be described with [Median & Range] because it is [a Skewed Continuous] variable

Question 7

40.0 | 94.9 | 19.7 to 29.6 | 5.0% | 64 | 23.5 to 26.2 | 2 | 5.1 | mean age=24.7 | £30001 and £40000 | median age = 24.5 | 48 | range=9.90 | up to £10000 | £40001 to £50000 | 6 | 53.3 | sd=1.95 | 10.2 | 5.1% | 40.7 | 54.2 | £20001 to £30000 | £10001 to £20000 | 89.8 | more than | 1.7 | 10.0

Check the remaining variables for any typos and delete them. In the 'clean' dataset, please describe the variables for gender, age and income.

The age of the individuals in our sample ranged from [19.7 to 29.6] years old with [mean age=24.7] ([sd=1.95])years.

In our sample, there were [64] individuals who identified themselves as females ([53.3]%), [48] individuals who identified themselves as males ([40.0]%), and [6] individuals who chose "other" as their gender identity ([5.0%]). There were [2] individuals who did not respond to this question ([1.7]%).

Most individuals earned between [£30001 and £40000] a year. Half of the individuals earned [more than] £30001 a year. [10.2]% of those who responded earned more than £60000.