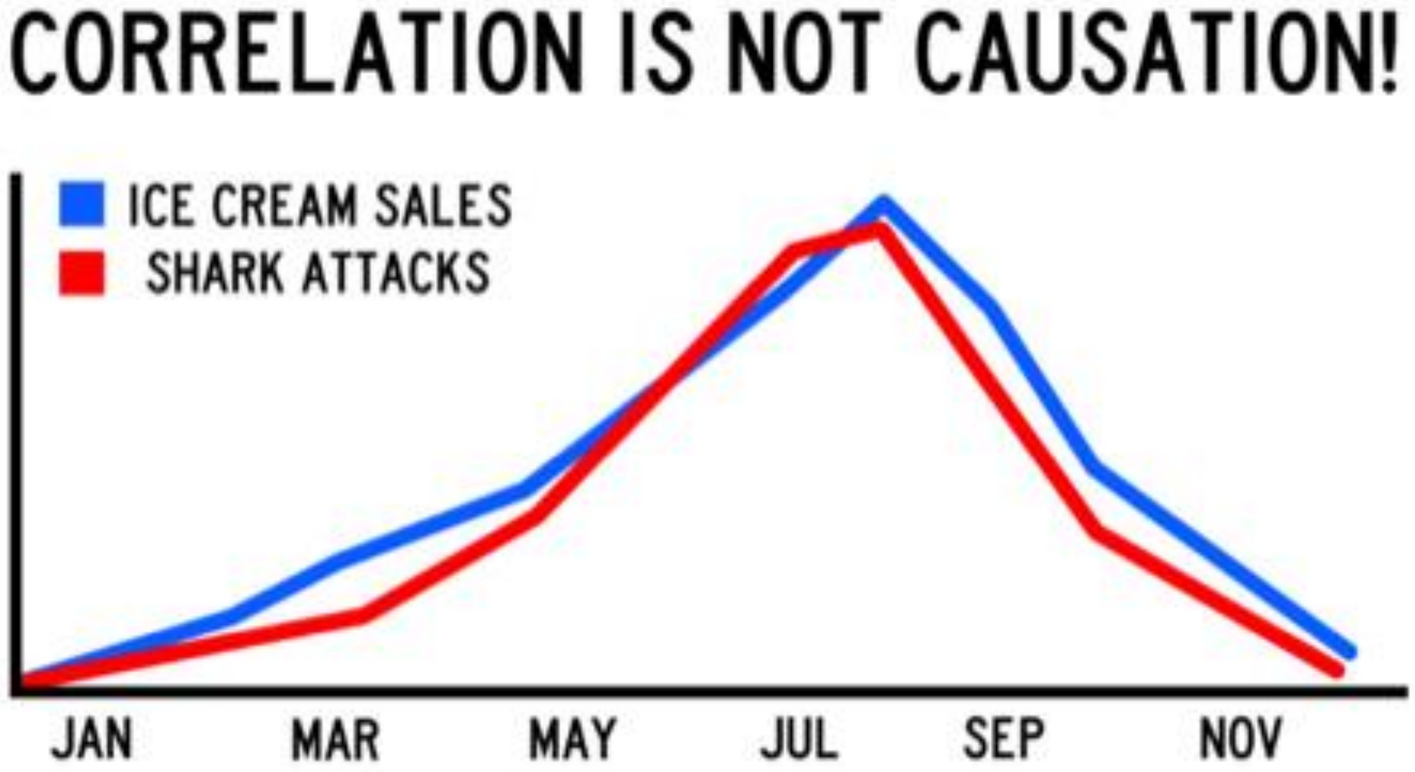


Intro & Review

ENVS225 Exploring the Social World

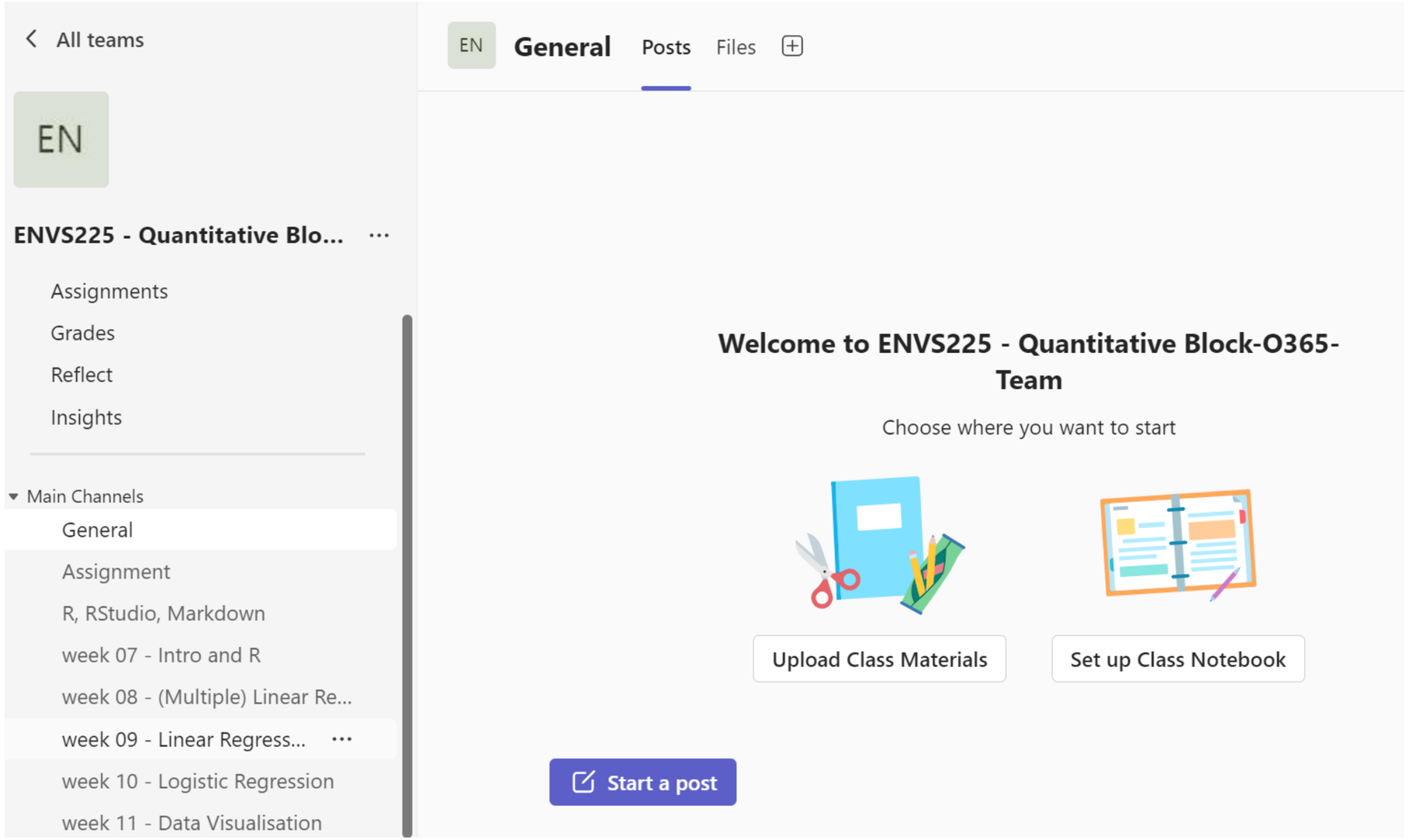
Gabriele Filomena
gfilo@liverpool.ac.uk

Quantitative Data Block: General Info

Lecture + Lab Sessions:
Friday 11 – 1 pm

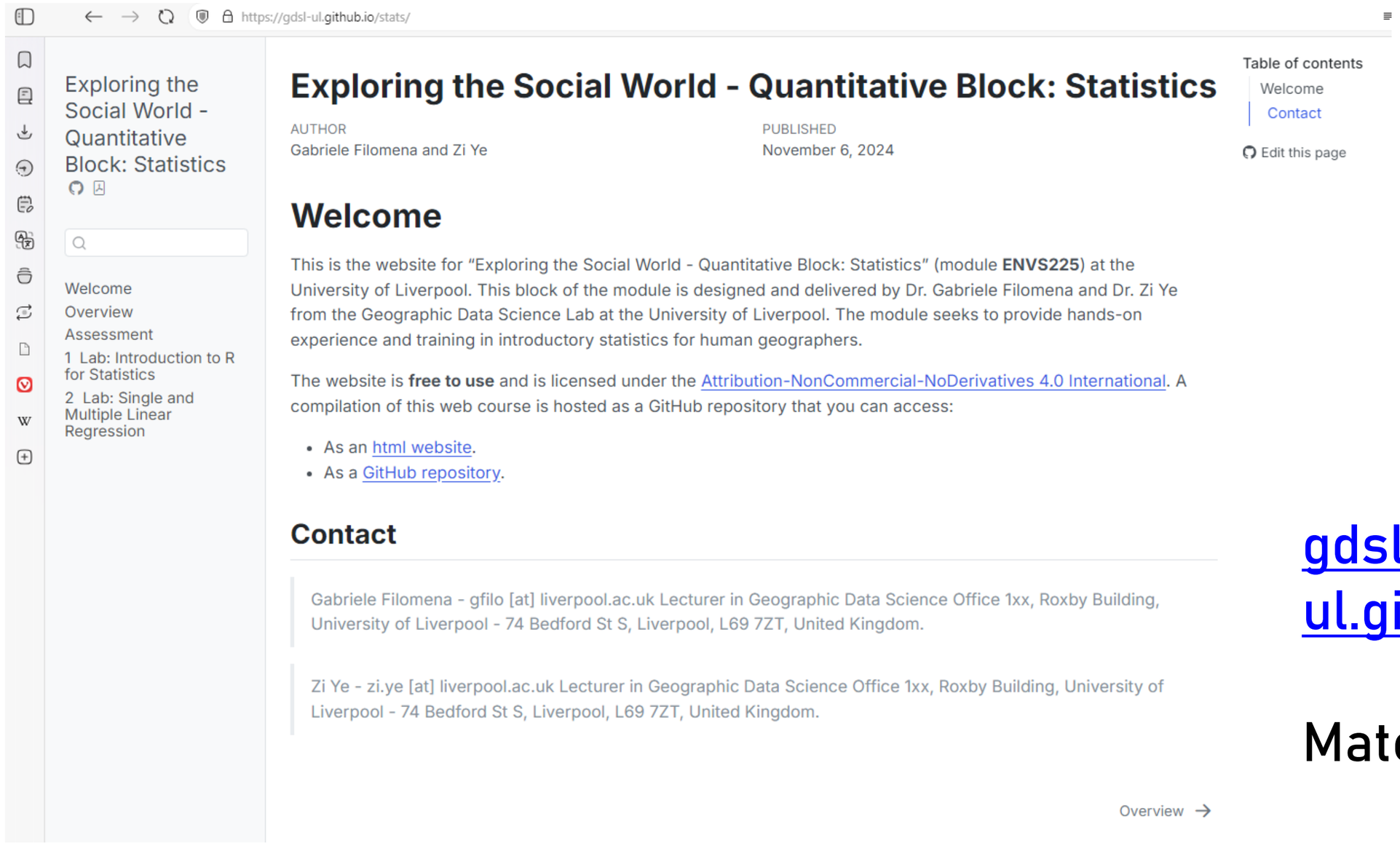
Week 8 -> *Central Teaching Hub, PCTC*

Quantitative Data Block: Questions -> MS Teams



[Join the team]

Quantitative Data Block: Webpage



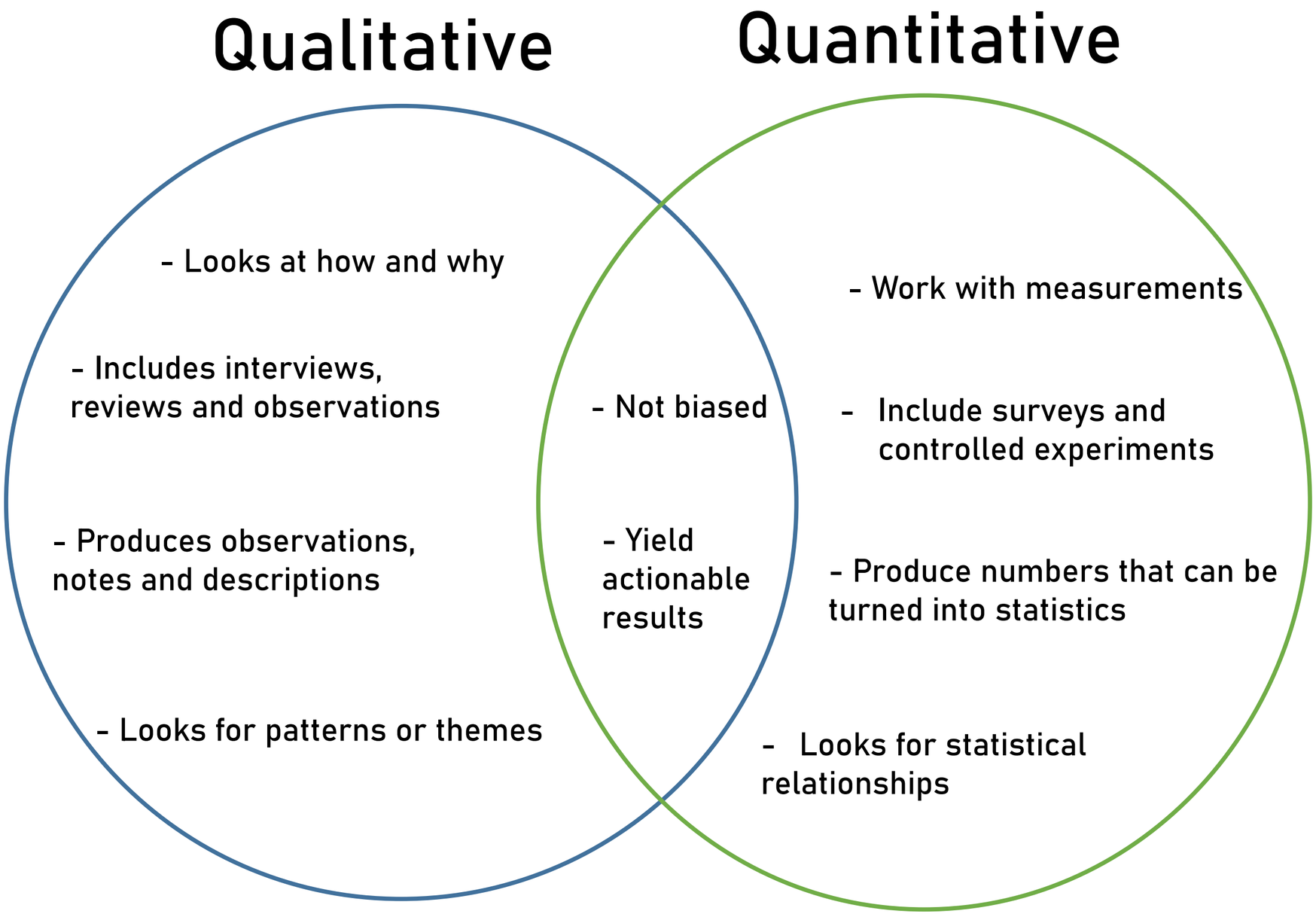
gdsl-ul.github.io/stats/

Material and Labs

Quantitative Data Block: Assignment

Writing a research report using one of the regression techniques learned during the module. The report should have the following sections: an introduction, a brief literature review, methods and data, results and/or discussion, and discussion or/and conclusion.

See webpage


Qualitative
Quantitative
- Looks at how and why
- Includes interviews, reviews and observations
- Produces observations, notes and descriptions
- Looks for patterns or themes
- Not biased
- Yield actionable results
- Work with measurements
- Include surveys and controlled experiments
- Produce numbers that can be turned into statistics
- Looks for statistical relationships

Why Both?

- Jobs
- Holistic understanding of problems
- Dissertation
- Better appreciation of what Geographers do

Learning goals

- Understand how to explore a dataset, containing a number of observations described by a set of variables.
- Demonstrate an understanding in the application and interpretation of commonly used quantitative research methods.
- Demonstrate an understanding of how to work with quantitative data to address real-world research questions.

Structure

7	**Introduction & Review**	GF
8	Single & Multiple Linear Regression	GF
9	Multiple Linear Regression with Categorical Variables	ZY
10	Logistic Regression	ZY
11	Data Visualisation	GF
12	Summary and Assessment Support	ZY

What are we going to work with?

nrKilled	nrWound	year	country	group	attack	target	weapon
0	0	2005	United Kingdom	Abu Hafs al-Masri Brigades	Bombing/Explosion	Transportation	Explosives/Bombs/Dynamite
0	0	2005	United Kingdom	Abu Hafs al-Masri Brigades	Bombing/Explosion	Transportation	Explosives/Bombs/Dynamite
0	0	2005	United Kingdom	Abu Hafs al-Masri Brigades	Bombing/Explosion	Transportation	Explosives/Bombs/Dynamite
0	0	2005	United Kingdom	Abu Hafs al-Masri Brigades	Bombing/Explosion	Transportation	Explosives/Bombs/Dynamite
0	1	1982	United Kingdom	Abu Nidal Organization (ANO)	Assassination	Government (Diplomatic)	Firearms
0	0	2014	United Kingdom	Anarchists	Facility/Infrastructure Attack	Business	Incendiary
0	0	2014	United Kingdom	Anarchists	Facility/Infrastructure Attack	Business	Incendiary
0	0	2014	United Kingdom	Anarchists	Facility/Infrastructure Attack	Business	Incendiary
0	0	2014	United Kingdom	Anarchists	Facility/Infrastructure Attack	Private Citizens & Property	Incendiary
0	0	2014	United Kingdom	Anarchists	Facility/Infrastructure Attack	Police	Incendiary
0	0	1984	United Kingdom	Animal Liberation Front (ALF)	Unarmed Assault	Private Citizens & Property	Chemical
0	0	1988	United Kingdom	Animal Liberation Front (ALF)	Facility/Infrastructure Attack	Business	Unknown
0	0	1988	United Kingdom	Animal Liberation Front (ALF)	Facility/Infrastructure Attack	Business	Unknown
0	0	1988	United Kingdom	Animal Liberation Front (ALF)	Facility/Infrastructure Attack	Business	Incendiary
0	0	1988	United Kingdom	Animal Liberation Front (ALF)	Facility/Infrastructure Attack	Business	Incendiary
0	0	1988	United Kingdom	Animal Liberation Front (ALF)	Facility/Infrastructure Attack	Business	Incendiary
0	0	1988	United Kingdom	Animal Liberation Front (ALF)	Facility/Infrastructure Attack	Business	Incendiary

Descriptive Statistics

Descriptive statistics

- Help us summarize the characteristics of our data/observations
- They form the initial step in a statistical analysis
- We get information about the centrality (e.g., mean, median)
- We get information about the variability (e.g., variance, standard deviation)

Mean

- *Mean* is another word for average.
- Most commonly used statistic that tells us about the centre of a data set.
- It is the average of the numbers in your dataset.

The sum of all values divided by the number of observations.

Median

- It is the middle point of your ordered data.
- If one orders the observations:

  *1,3,**5**,8,10*

  The median is the number in the middle

- When the observations are even:

  *1,3,**5**,6,8,10*

  Average the mid values -> 5.5

Difference between Mean and Median

- Both are single values that represent the center of a data set.
- But median is more resilient to outliers.

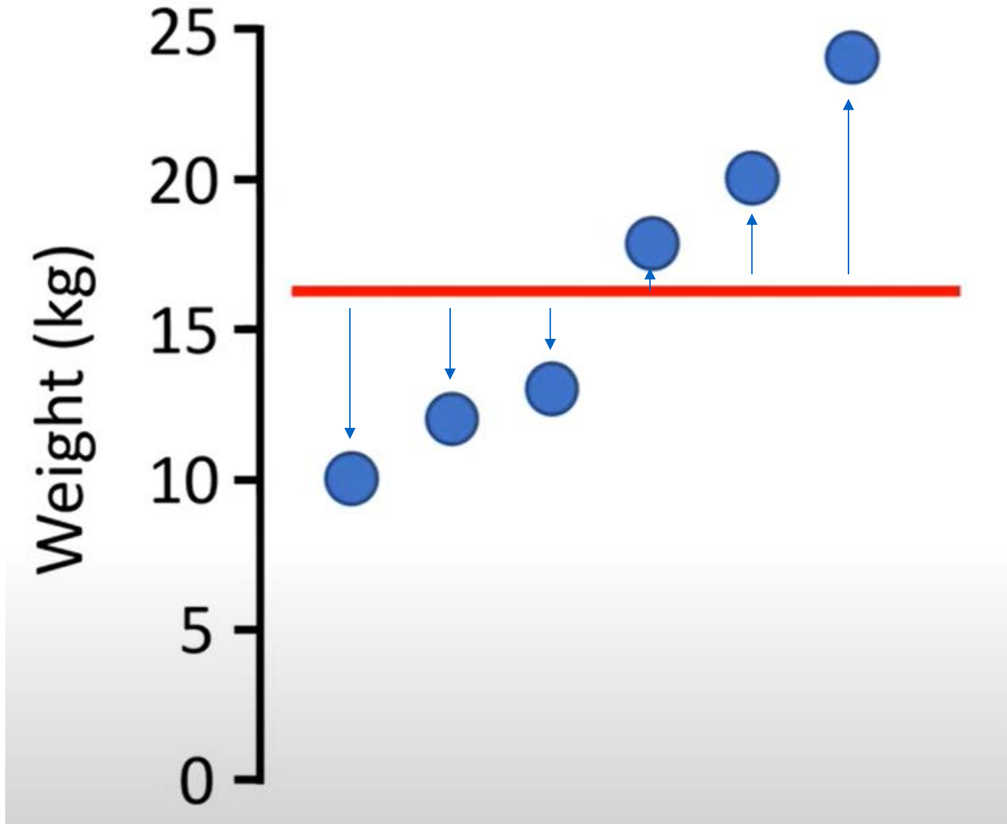
E.g.

$$3,1,4,2.5,50 \quad (n=5)$$

- *Mean*: $\frac{3+1+4+2.5+50}{5}=12.1$
- *Median*: $1, 2.5, 3, 4, 50 = 3$

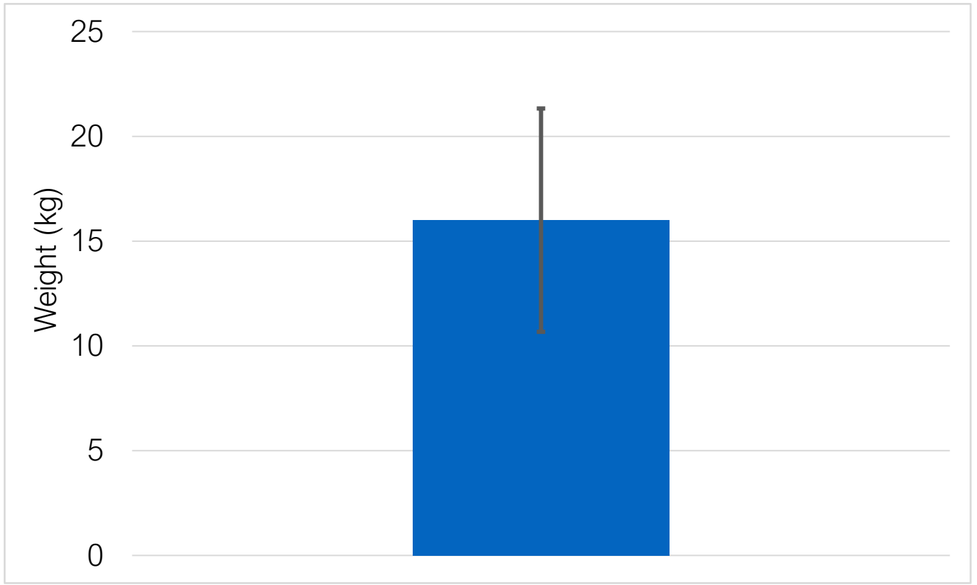
Variance (s^2)

- A measure to show how spread out our observations are
- Based on the distance of each observation from the mean
- The average of the squared differences from the mean.



Standard deviation (*s*)

- The square root of *variance*
- The result is in the same measurement units as our data (e.g., kg).
- This makes it easier to interpret and understand the spread of the data than *variance* alone.



- *Small value indicates data are gathered close to the mean*
- *Large value indicates data are gathered far from the mean*

Correlation

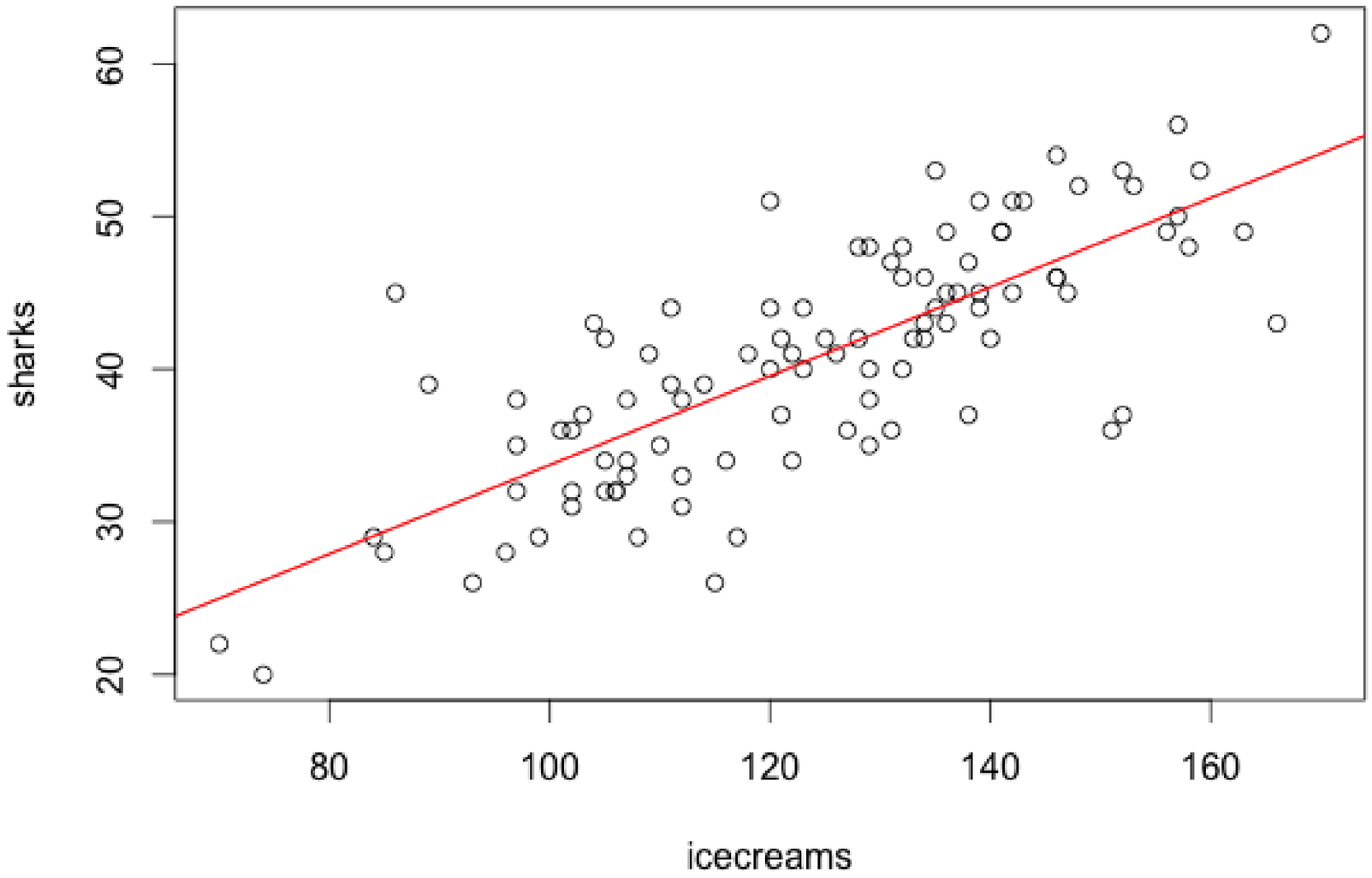
- Describes an association between variables.
- When one variable changes, so does the other.
- Covariation, rather than a direct causal link.

Examples:

- Ice cream sales and shark attacks.
- Time spent in shops and temperature.

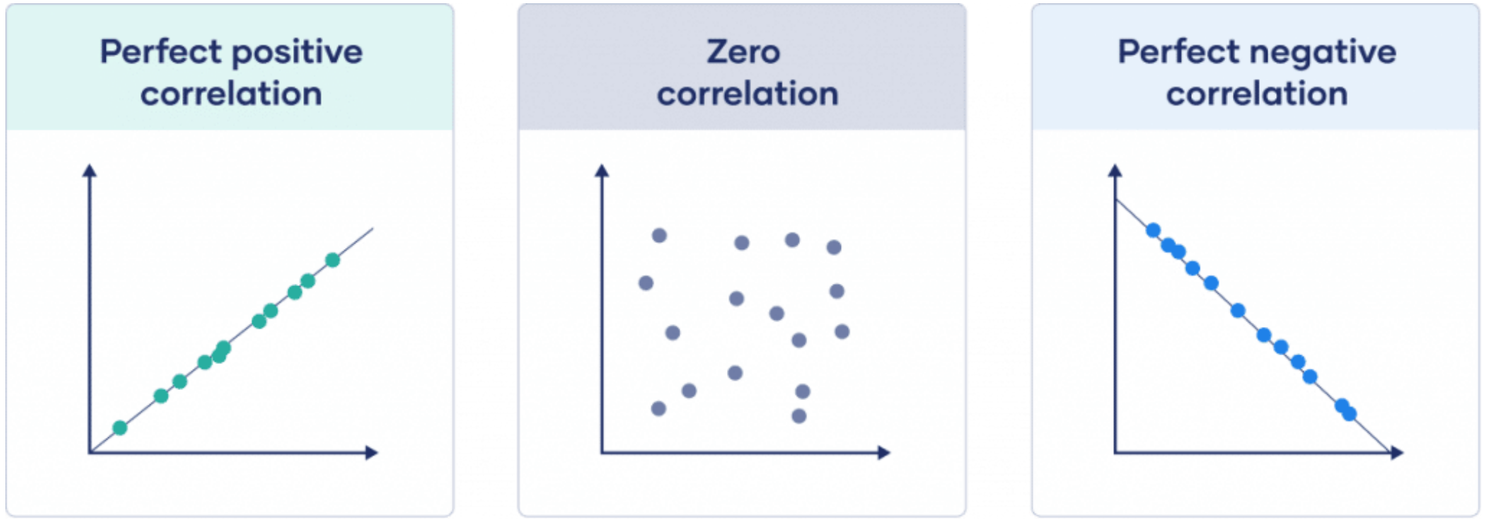
Associations between Variables

What is Correlation?



Pearson Correlation

The Pearson correlation coefficient (*r*) measures the strength and direction of the relationship between Continuous variables.



r = 1

Same direction

r = 0

r = -1

Opposite direction

Single Linear Regression

- Regression is used to study the relationship between two variables.
- How a change in one variable (e.g., someone's exercise habits) can predict the outcome of another variable (e.g., general health).

DEPENDENT <- INDEPENDENT VARIABLES

=

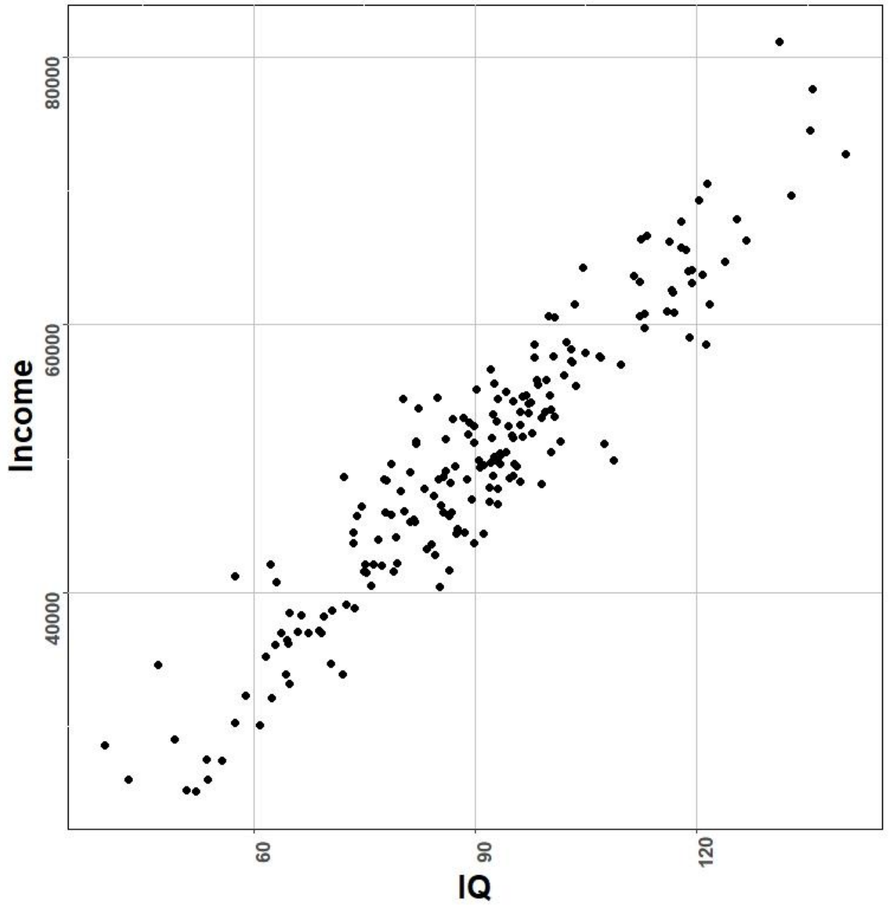
OUTCOME/RESPONSE <- PREDICTOR/EXPLANATORY

Examples

- IQ affects income (IQ is the IV and income is the DV)?
- Study time affects grades (hours of study time is the IV and average grade is the DV)?
- Exercise affects blood pressure (hours of exercise is the IV and blood pressure is the DV)?

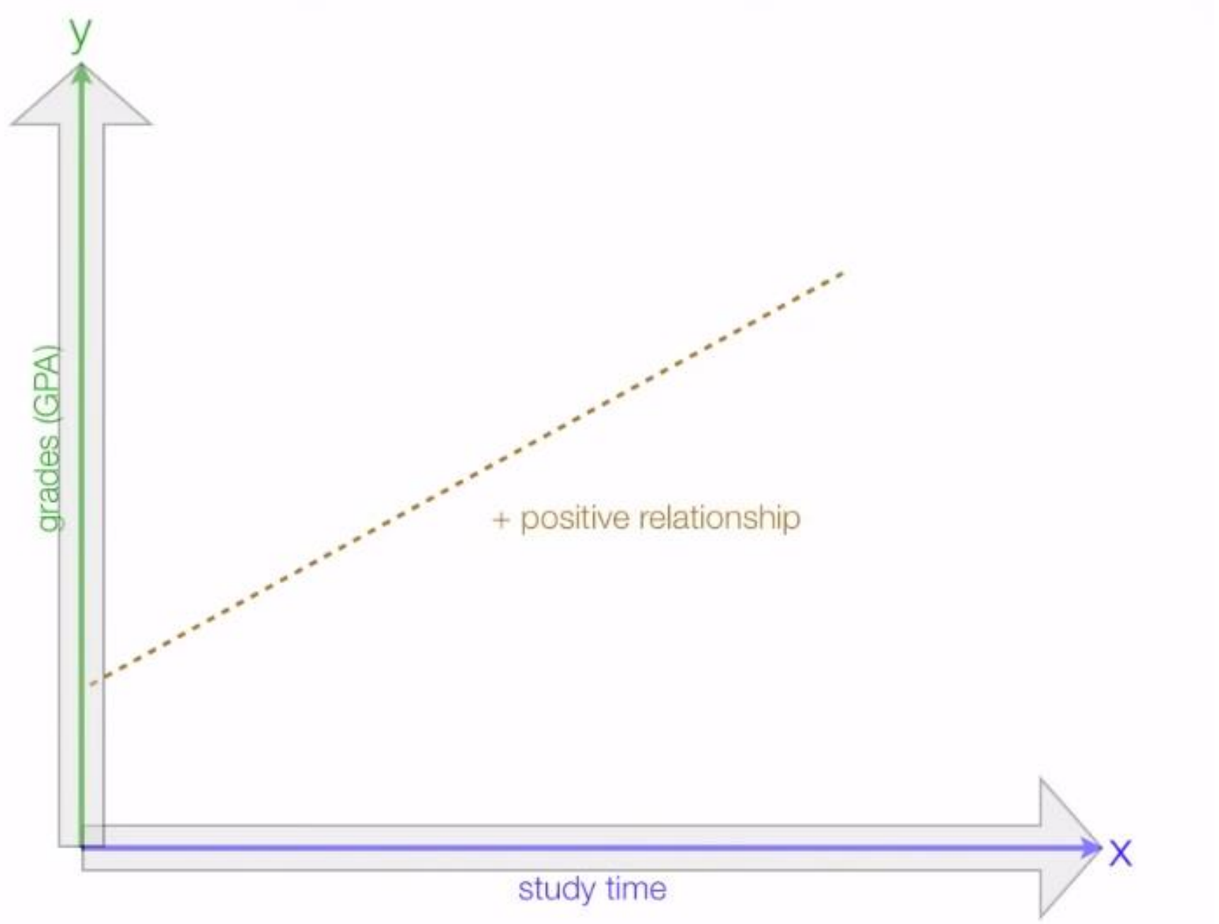
Displaying the data

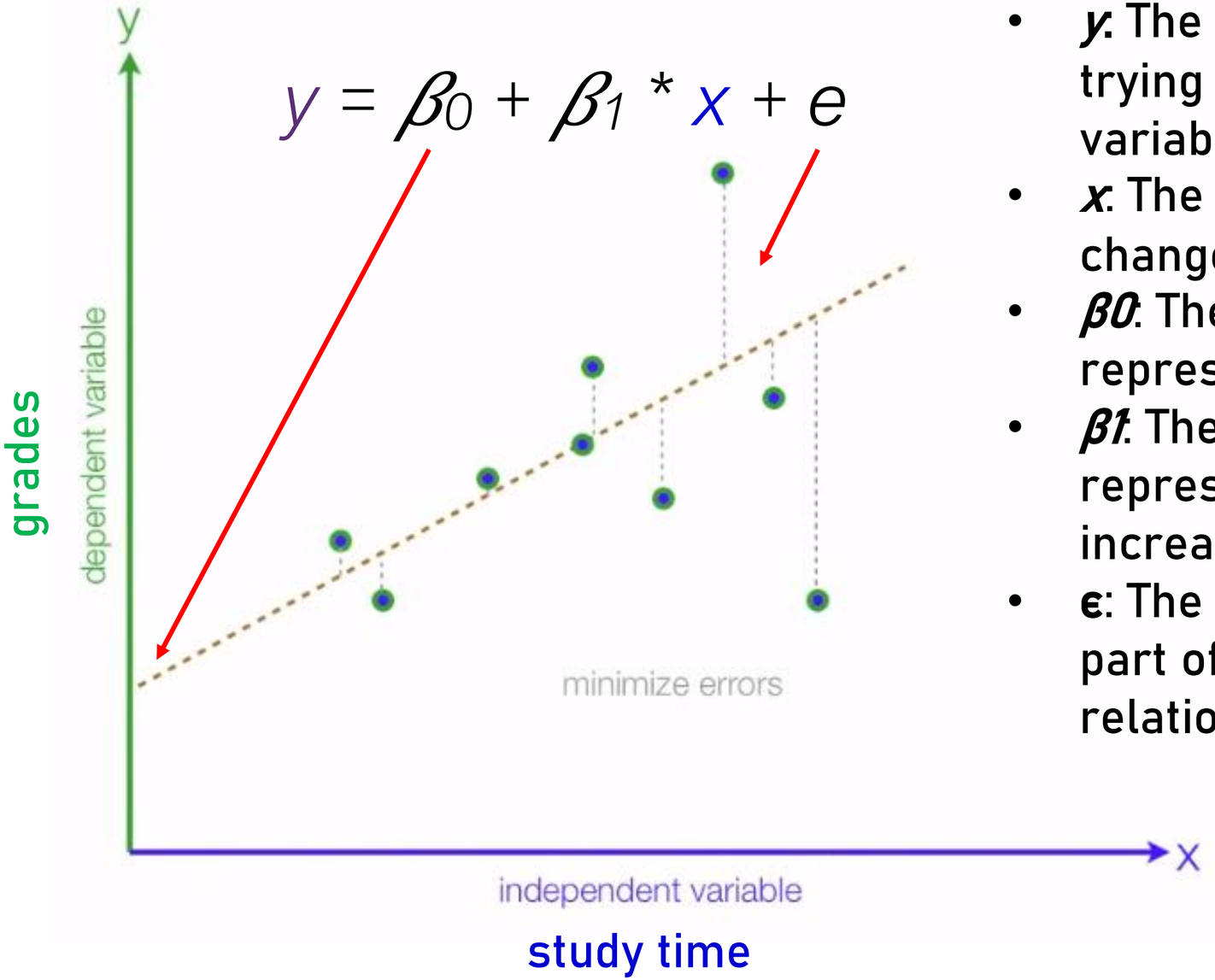
When both the DV and IV are numerical, we can represent data in the form of a scatterplot.



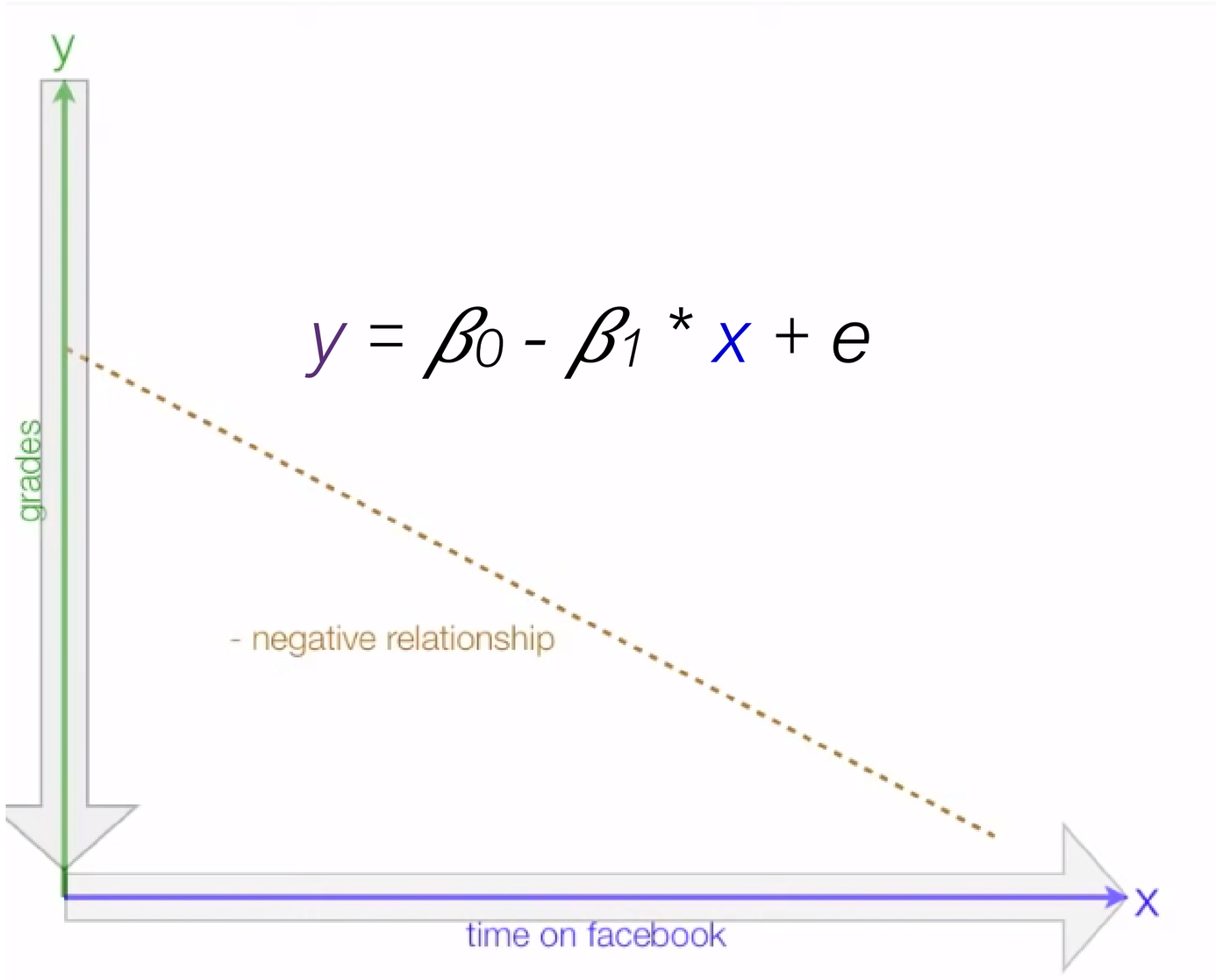
Single Linear Regression

- Straight line prediction model.
- As an independent variable changes, what happens to the dependent variable?





- ***y***: The dependent variable. The outcome we are trying to explain based on the independent variable.
- ***x***: The independent variable used to explain changes in the dependent variable.
- ***β0***: The intercept of the regression line. It represents the expected value of *y* when *x* is 0.
- ***β1***: The slope of the regression line. It represents the change in *y* for a one-unit increase in *x*.
- **ϵ**: The error term (or residual). It represents the part of *y* that cannot be explained by the linear relationship with *x*.



$$y = \beta_0 - \beta_1 * x + e$$

Correlation vs Single Linear Regression

- Correlation quantifies the strength and direction of the linear relationship between two variables but doesn't give information about the equation of the line.

- Single Linear Regression goes further by providing an equation that describes the linear relationship.

- Single Linear Regression models the relationship between two variables.

Week 7 Task

- Read the paper *Defining Research Questions* on Canvas.
- Complete today's practice.
- Familiarise yourself with RStudio.