Class 31

DATA1220-55, Fall 2024

Sarah E. Grabinski

2024-11-20

Review: 2 Numeric Variables

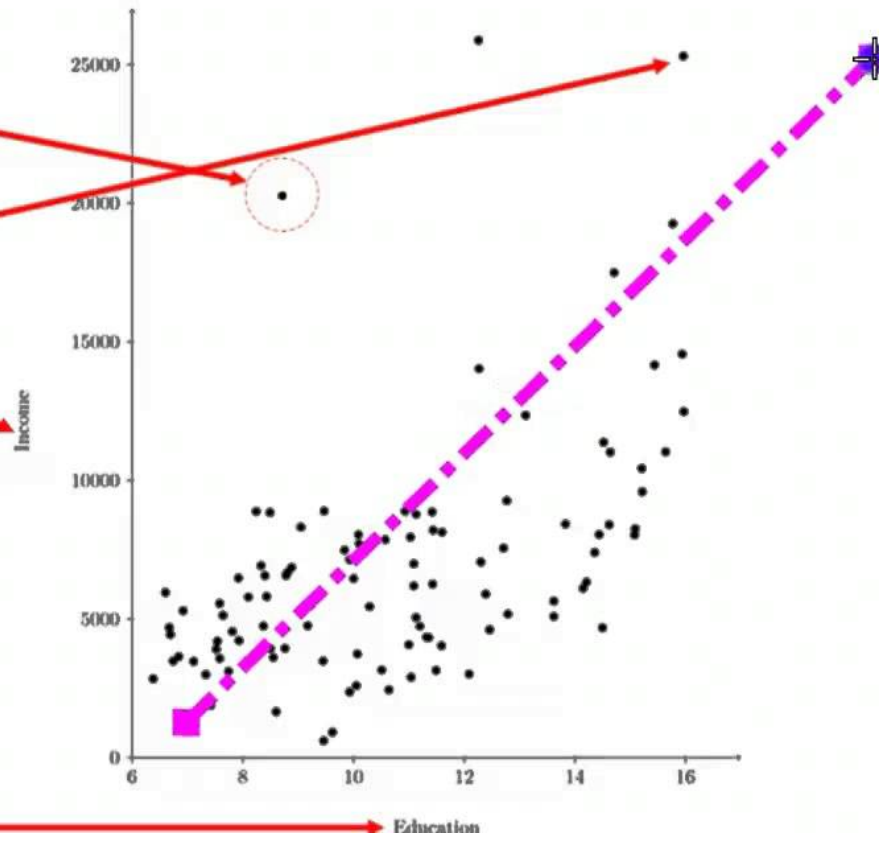
Understand how to read a scatter plot.

Outlier: a point that is far away from the pattern of the other data points

Influential point: a point that is far away from the other data points

Response variable / dependent variable: the variable that reacts to the explanatory variable; it is plotted on the y-axis

Explanatory variable / independent variable: the variable that causes the other variable to change; it is plotted on the x-axis



Review: Describing Associations

- ***Independence***: an increase in X is not associated with a change in Y

Review: Describing Associations

- ***Independence***: an increase in X is not associated with a change in Y
- ***Positive association***: an increase in X is associated with an *increase* in Y

Review: Describing Associations

- ***Independence***: an increase in X is not associated with a change in Y
- ***Positive association***: an increase in X is associated with an *increase* in Y
- ***Negative association***: an increase in X is associated with a *decrease* in Y

Review: Describing Associations

- ***Independence***: an increase in X is not associated with a change in Y
- ***Positive association***: an increase in X is associated with an *increase* in Y
- ***Negative association***: an increase in X is associated with a *decrease* in Y
- ***Weak association***: data points are very far apart from each other

Review: Describing Associations

- ***Independence***: an increase in X is not associated with a change in Y
- ***Positive association***: an increase in X is associated with an *increase* in Y
- ***Negative association***: an increase in X is associated with a *decrease* in Y
- ***Weak association***: data points are very far apart from each other
- ***Strong association***: data points are tightly clustered

Pratice

Which image shows a ***positive*** relationship between the explanatory and response variables?

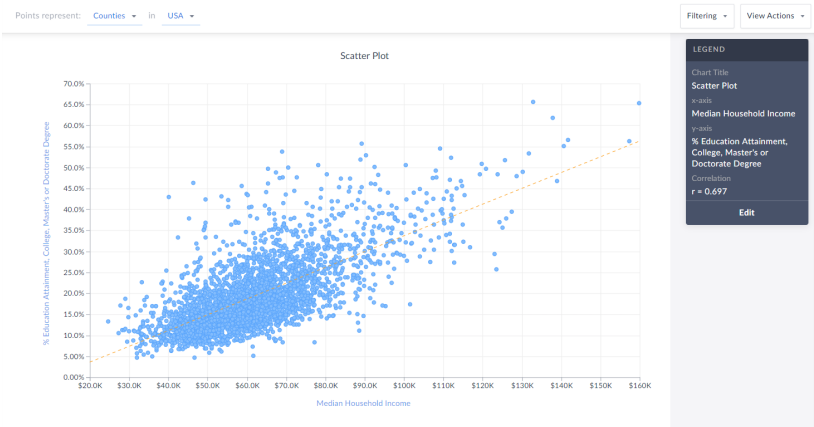


Figure 1: Income vs Education

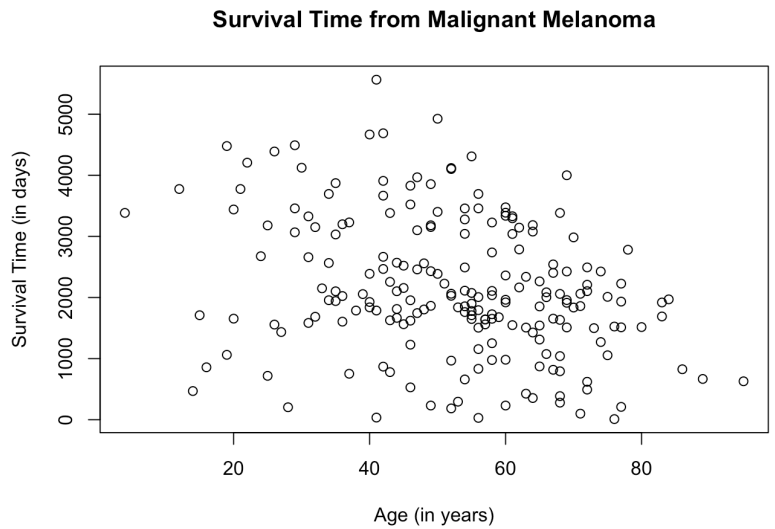


Figure 2: Age vs Survival

Practice

Which image shows a ***weak*** relationship between the explanatory and response variables?

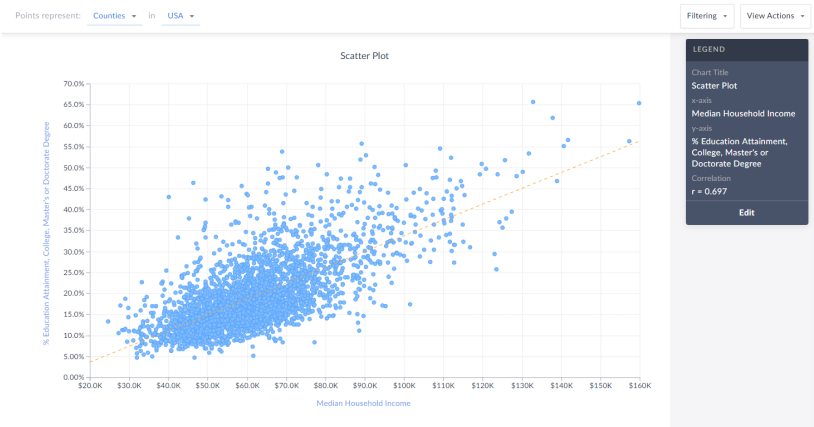


Figure 3: Income vs Education

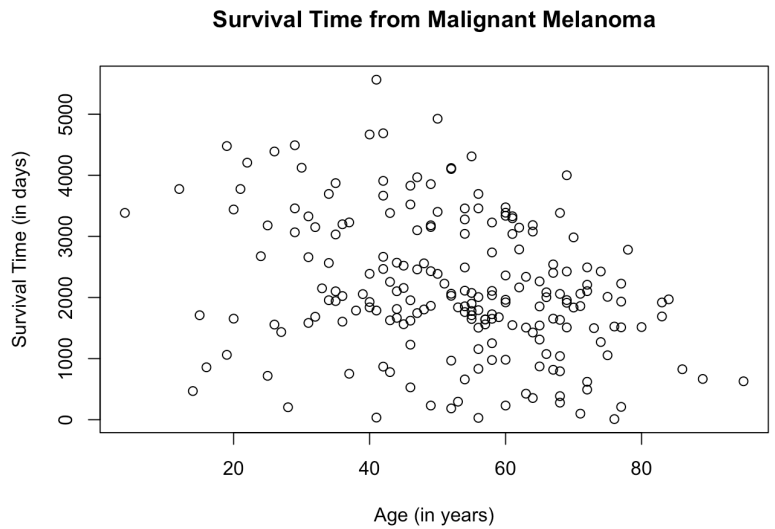


Figure 4: Age vs Survival

Correlation

- Describes the direction and strength of the association between 2 numeric variables

Correlation

- Describes the direction and strength of the association between 2 numeric variables
- A correlation ranges from -1 to 1
  - A perfect negative correlation equals -1
  - A perfect positive correlation equals 1

Correlation

- Describes the direction and strength of the association between 2 numeric variables
- A correlation ranges from -1 to 1
  - A perfect negative correlation equals -1
  - A perfect positive correlation equals 1
- A correlation of 0 indicates the two variables are independent (no relationship)

Correlation

- Describes the direction and strength of the association between 2 numeric variables
- A correlation ranges from -1 to 1
  - A perfect negative correlation equals -1
  - A perfect positive correlation equals 1
- A correlation of 0 indicates the two variables are independent (no relationship)
- We use the Pearson correlation for linear relationships

Linear vs Non-Linear