CHANGES TODO

Print graphs all on a single page... collating was a nightmare

Cut out data exercise for next year... not enough time

Notes

5: Reflection on working styles/practices for learning R. Share tips.

10: What is the point of repeated measures? What benefits do we gain?

15: Generate examples of repeated measures data.

20: School exercise

30: Schools feedback/discussion

35: More schools, less information plot exercise

40: Quick recap on mean and variance.... Sum of the squares. Std Deviation is sqrt of variance. (Sample variance is N-1)

50: Independent and identically distributed assumption. Ask them to generate examples which would/would not be IID

60: Break

70: Splitting the error graphs ... ask students to guess meaning

95: Explictly go through splitting error term slide and intro to mixed models

105: RCT example: Followup question: Given a fixed budget, would it become more or less important to sample as many different patients as possible, compared with sampling the same patients repeatedly?

115: Quick demo of reshaping and plotting the fit data.

Homework: Exercise online - reshaping data in prep for next week

Schools notes

- Applied perspective -> no single right answer
- wouldn't pick one of the answers near the extremes because
- balancing different concerns:

These:

- Generalisability to all schools
- Practical/cost/time (although strictly this wasn't part of the question)
- Fair estimate for each school (because low scores could have consequences for teachers, if results were published)
- See if schools *vary* in how effective they are?

But this question was about estimating the **average effectiveness** across the whole of the UK. So it's really a question about how to maximise **power**.

Highlight that POWER is not always highest with highest N.... also have to think about k (groups)

If students within schools are similar then ***might gain more information*** when we sample from a new school, even though that costs us more and the total N goes down.

Schools

Imagine you work for the UK Department for Education (DfE).

Your task is to find out whether a new educational intervention is effective. The new intervention has been rolled out in 1000 randomly-selected schools in the UK.

- You have a budget of around £10,000.
- The cost to visit a single school is £100
- The cost to measure each pupil is £10
- The average size of a school in the UK is roughly 250.

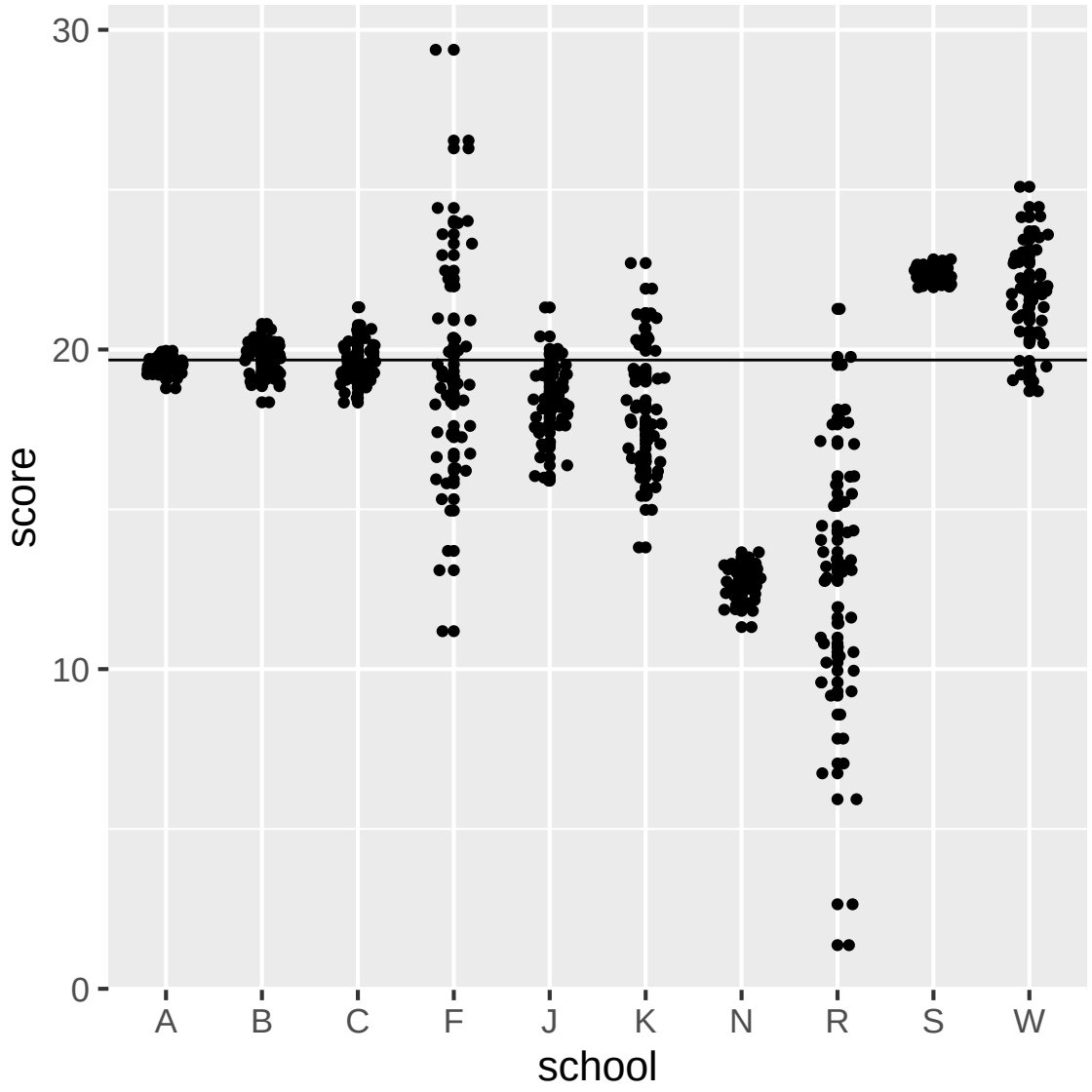
Would you prefer to collect data from:

- 970 children across 3 schools
- 870 children across 13 schools
- 770 children across 23 schools
- 670 children across 33 schools
- 570 children across 43 schools
- 470 children across 53 schools
- 370 children across 63 schools
- 270 children across 73 schools
- 170 children across 83 schools

These options all cost around £10,000.

Discuss the problem as a small group. Identify the different tradeoffs involved.

Schools plot task

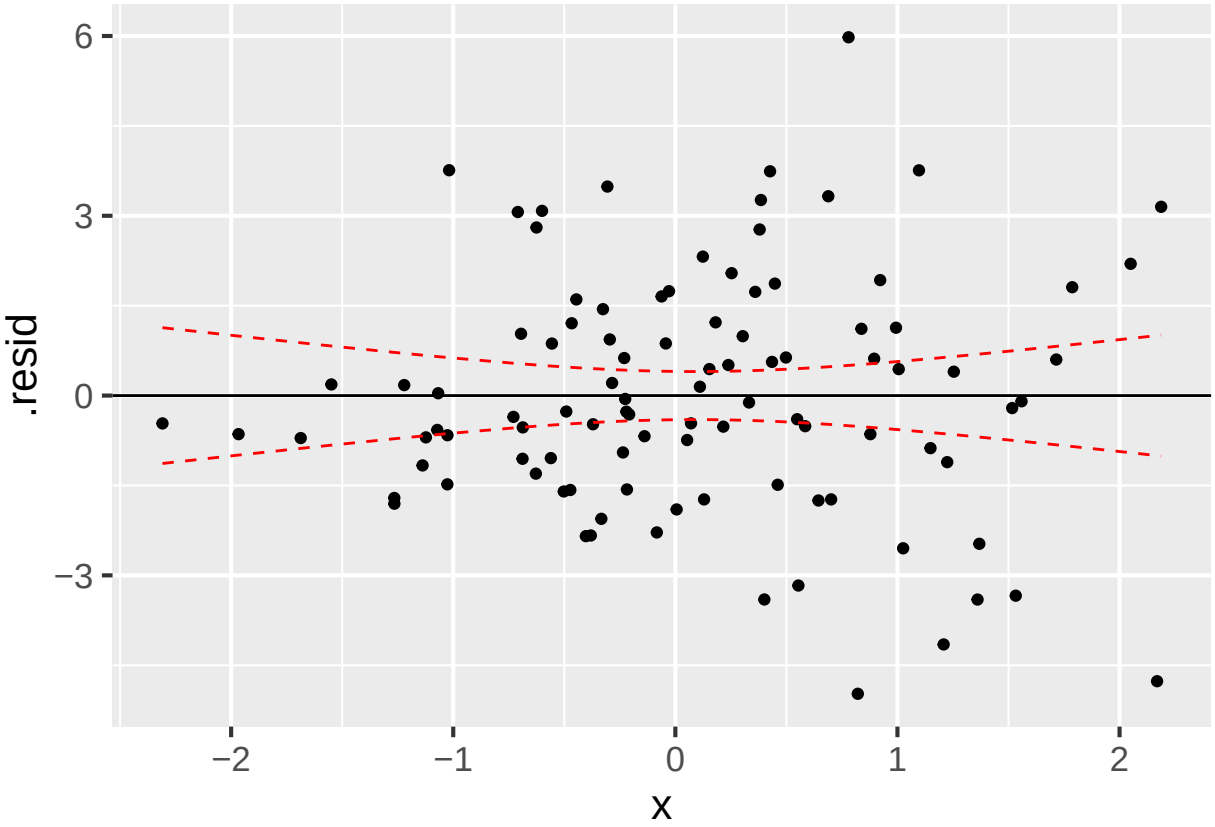


- Which schools would you most and least want to sample another child from? Why?
- Why would it be better to sample from another school, than another pupil from school in the set plotted?

IID plots

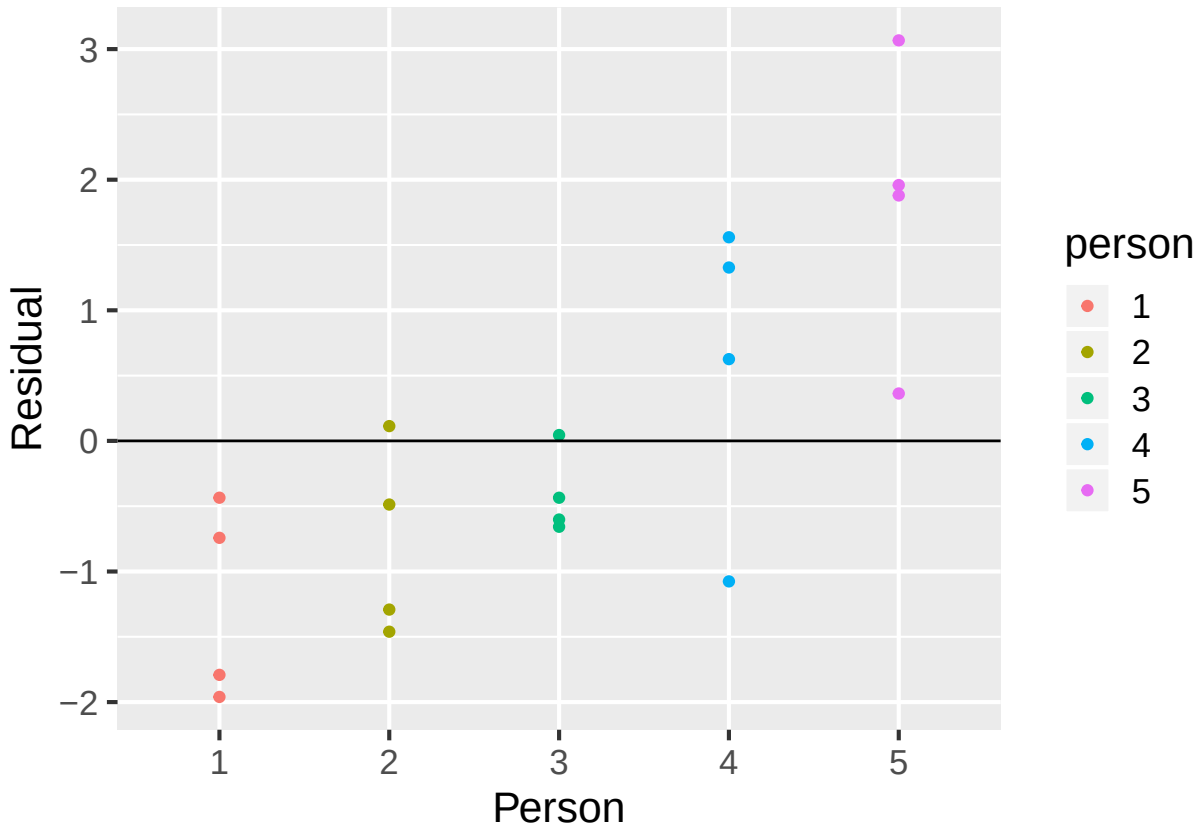
Q: Are residuals equally likely to fall outside 2xSE for all values of X?

```
Saving 6.5 x 4.5 in image
```



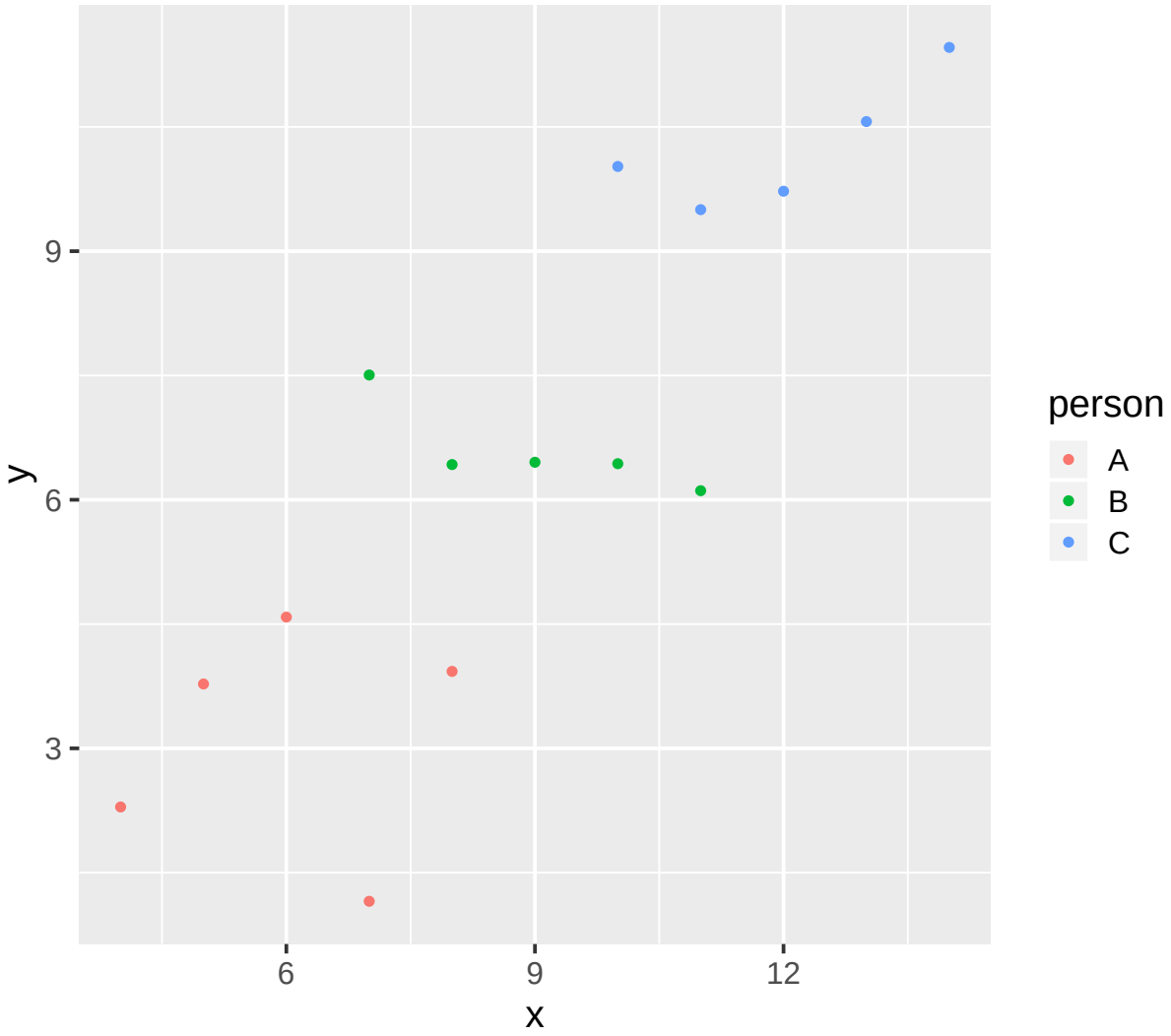
Q: Are residuals equally spaced for different people?

```
Saving 6.5 x 4.5 in image
```


Residual
3
2
1
0
−1
−2
1
2
3
4
5
Person
person
1
2
3
4
5

Label the plots exercise

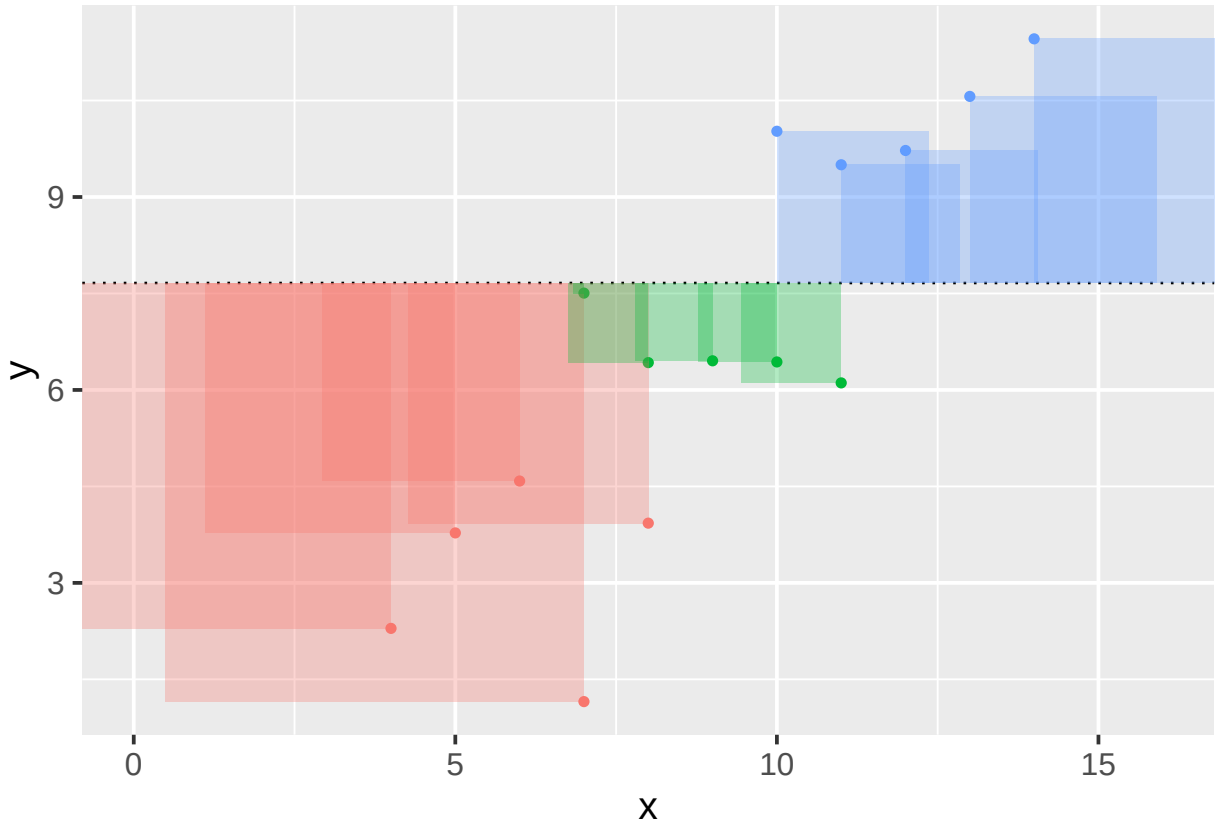
This plot shows 15 datapoints, for 3 individuals within a repeated measures study:

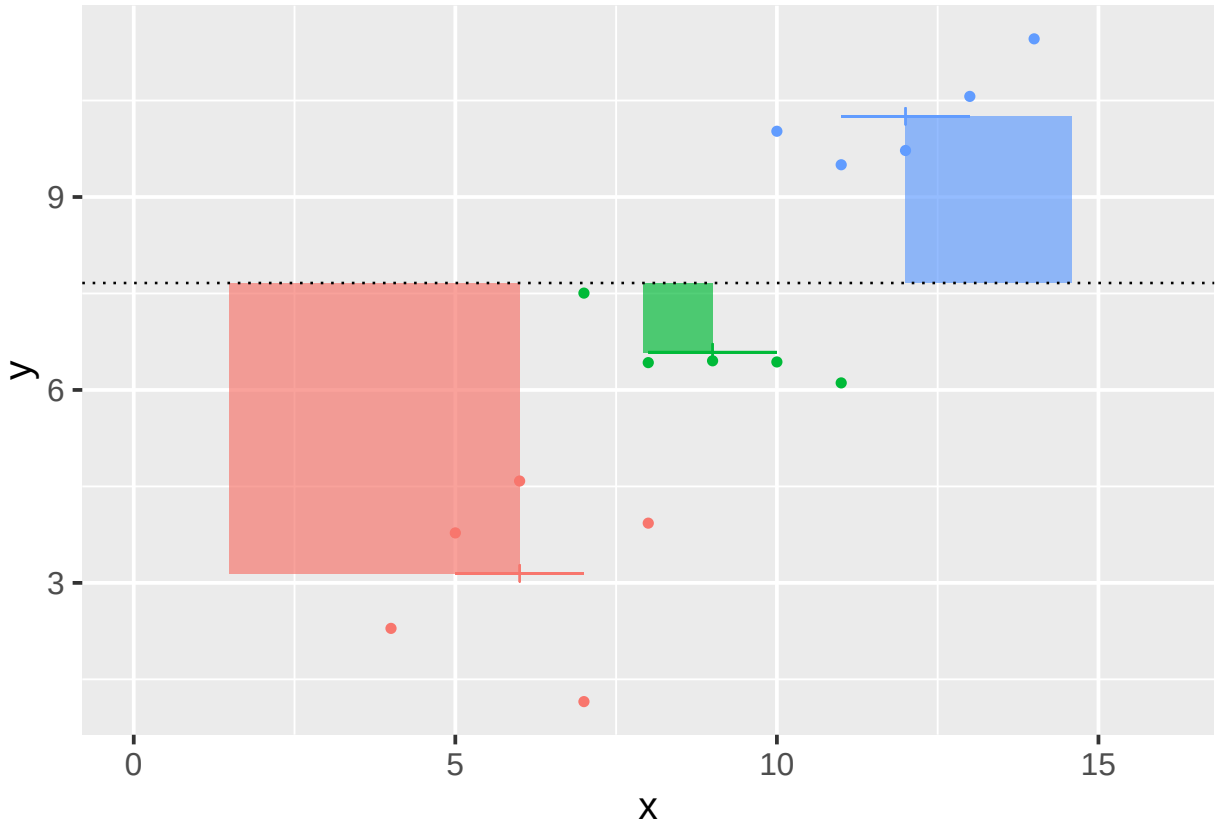


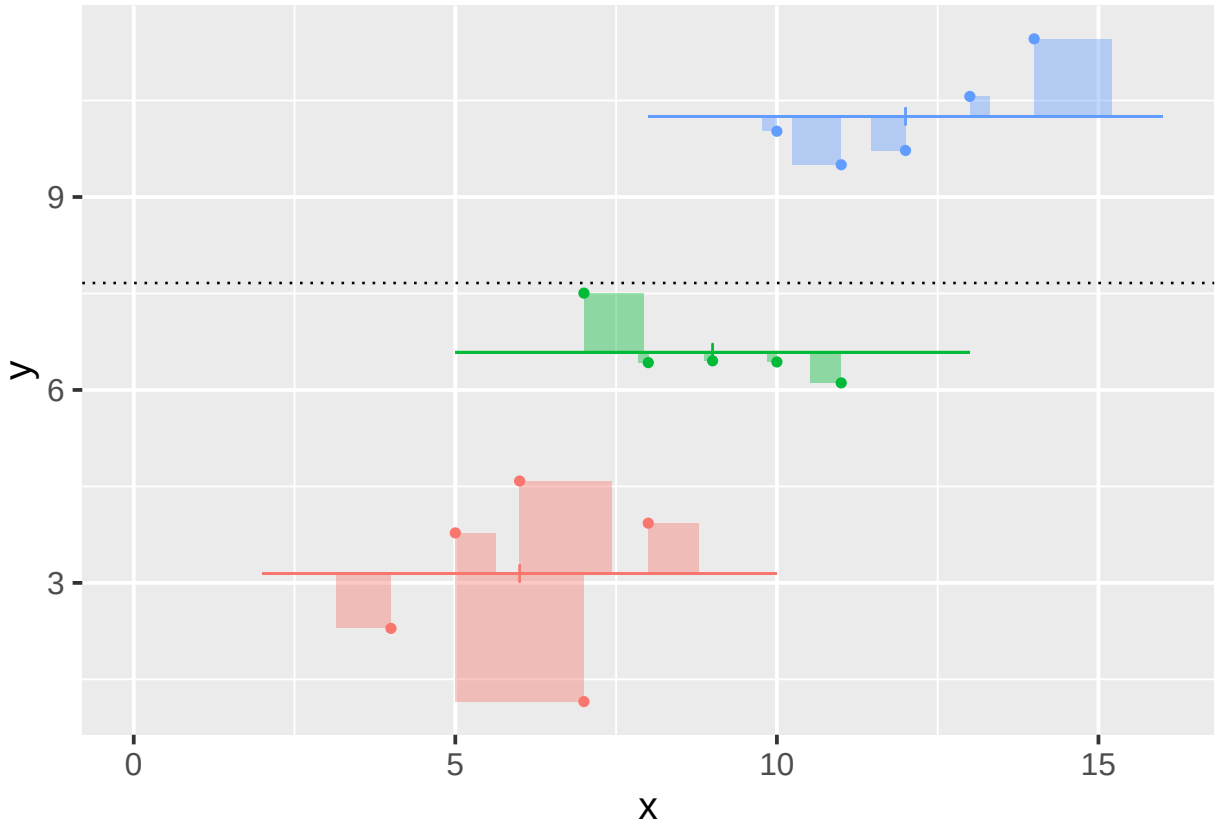
The x-axis could be anything; for this exercise it's not relevant, but is displayed to make the plot easier to read. If it helps, you can imagine it represents time of day.

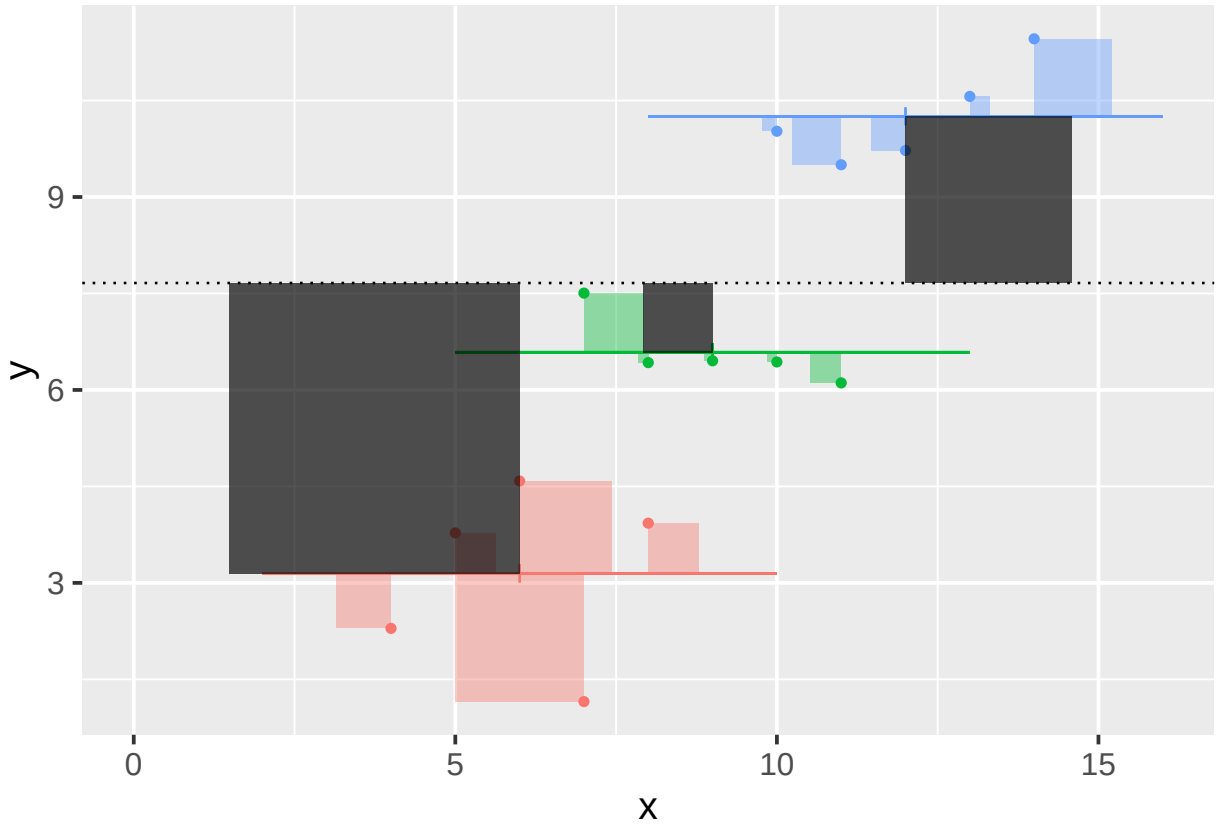
Your job is to label each of the following plots (which use the same data), and indicate which of these concepts they illustrate:

- Total variance
- Variation between individuals
- Variation within individuals
- The relative proportion of variance within and between individuals.


9
6
3
y
0
5
10
15
x


9
6
3
y
0
5
10
15
x


9
6
3
y
0
5
10
15
x

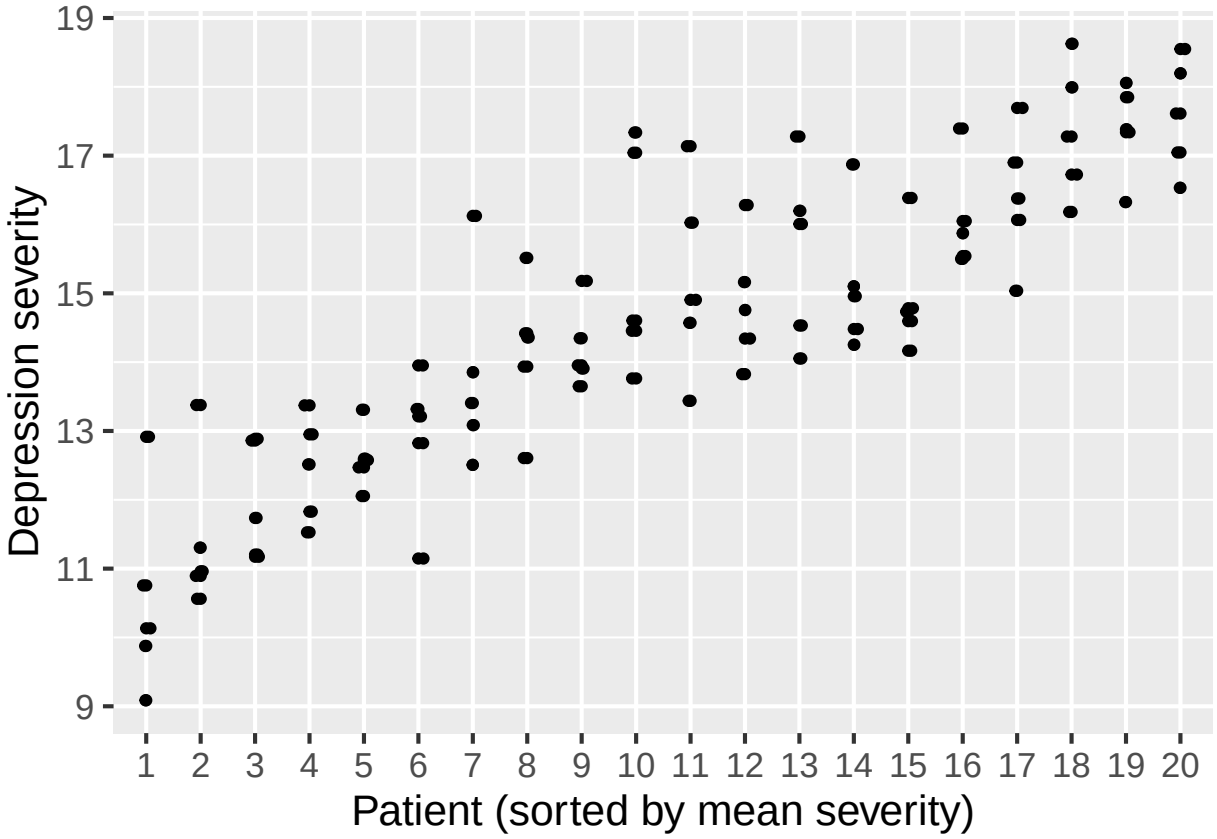

9
6
3
y
0
5
10
15
x

RCT Example

Imagine these are data from a clinical trial for a treatment for depression. We measured patients' depression in the year after they were treated.

The plot below shows 5 measurements taken from each person, during this year:

```
Saving 6.5 x 4.5 in image
```



In the first study we only included women, and excluded patients with any medical or psychological comorbidities.

Now imagine we re-run the study, but change the entry criteria to recruit a much wider group of patients.

- How would this plot change?
- Would the total amount of variation in depression increase or decrease?
- Would there be more variation *within* or *between* patients, or both? Which would see the greater increase?