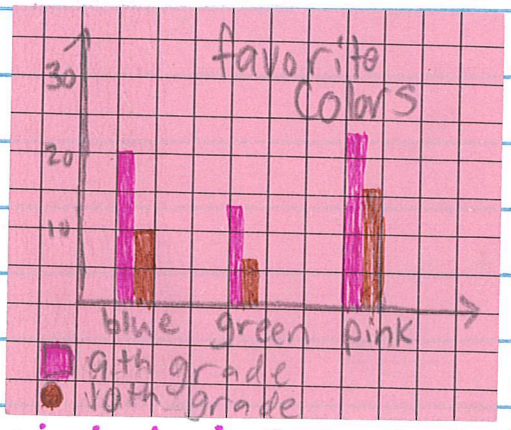
Unit 2 - Scatterplots, LSRL, Relative frequency, data graphs

Side by Side Graphs

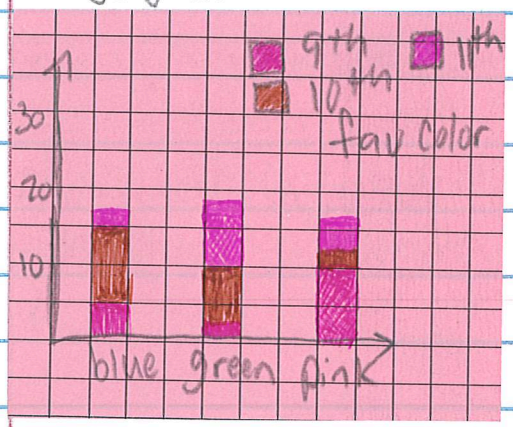
- Categorical, # values
- labels on sides
- next to each other
- swappable variables



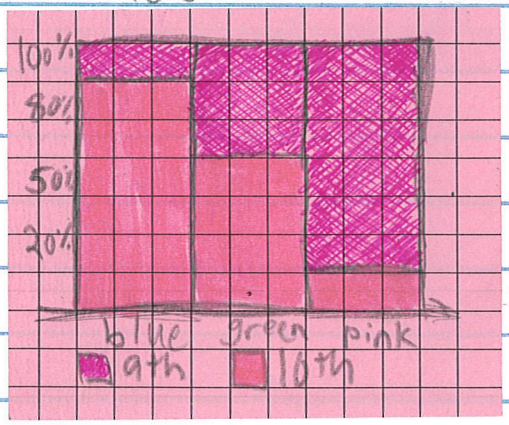
Segmented

- Swappable Variables
- 1st var on bottom, second as color on side

Segmented:



Mosaic:



Mosaic

- Uses relativ
frequencies to better compar
Variables.
- totals are not the same

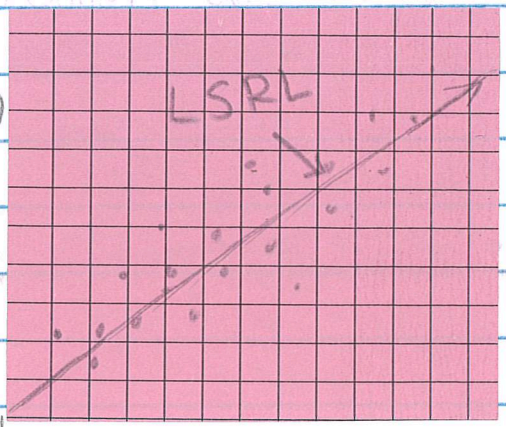
beware! The proportion of values may be higher, but the real value can be differen due to different totals.

Relative Frequency

- Also called Conditional frequency
- Probability of one event given another event must occur
- Value / column or row total
- In rel fre table, values must add to the totals for rows and columns.

Scatterplots

- Describe:
  - Direction (pos or neg)
  - unusual features: outliers, clusters, gaps
  - linear, non linear
  - strength: weak, med, strong. + Context



- Outliers
  - High leverage point
    - far from domain
    - doesn't affect residuals
    - close to LSRL
    - increases variance
  - Influential point
    - In domain
    - outside range
    - very influential on residuals
    - far from LSRL
    - A point can be both influential and high leverage

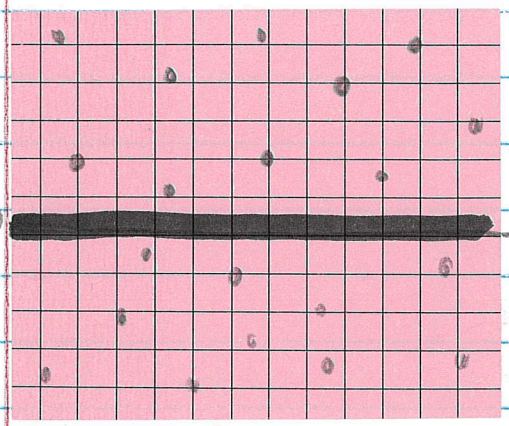
Correlation Coefficient (r) - How closely the scatterplot fits the least² Regline. From -1 to 1. -1 means perfect negative slope, 0 is no correlation, and 1 is perfect positive slope.

Coefficient of Determination (r^2) - What % of the respons variable is controlled by explainatory variable. Higher means x-var has more control over y-var. Smaller than r.

Residuals - How far observed value is from expected. Expected = plug in x value to LSRL. Ex) $y = 5.4(5) - 2$. Positive means underestimated, negative is overestimated. 0 = perfect.

Residual Graph

← linear model



- On x-axis is the observed value
- On y-axis is the residual value.
- A random pattern means linear is a good fit.
- Sres means the residuals are spread out by standard deviation of observed and residual values.

Calculator functions:

How to get residuals, expected, r^2, r, LSRL, Sres:

1. Enter x-values into L_1, y-values into L_2
2. To get **LSRL**, r^2, r: Stat, Calc, #4 LinReg
3. For **expected values**: go to L_3, enter $y=m(L_1)+b$ or the LSRL.
4. For **residual values**: go to L_4, enter L_2-L_3, or do obs-exp
5. For **Sres**: Calc, 1-var stats on L_4, Sx = Standard deviation of Res.

♡ Use this data to find out if the expected values are accurate or not, and if linear is a good fit.