RESULTS AND DISCUSSION SECTION WRITING

EID EMARY

MOTIVATION

- Many papers are rejected not because the research is weak, but because the authors:
  - Do not connect findings to the research question
  - Overstate conclusions
  - Ignore unexpected results
  - Repeat tables without insight
  - Many researchers confuse these two sections.
    - Results = What you found
    - Discussion = What the findings mean

THE RESULTS SECTION-MEANING

- Results = What you found in connection with the research questions.
- The results section should present data in a logical sequence without bias or interpretation, only state the facts.
- Findings may be reported in written text, tables, graphs, and other illustrations.
- Results are in a form of qualitative and/or quantitative measurements.

QUALITATIVE DATA

- **Qualitative** research uses non-numerical, descriptive data, such as language, themes, and ideas that reflect the human experience.
- It is often collected by using interviews, survey questionnaires, or archival research.
- Many disciplines that fall within the social sciences will use qualitative data.
- Example: to evaluate the output of an image generation system, rank the quality of the following image as {good, average, poor}

QUANTITATIVE DATA

- **Quantitative** research uses measurable (or quantifiable) numerical data that can be used to observe trends, make predictions, run experiments, and test hypotheses.
- Most disciplines that fall under STEM (Science, Technology, Engineering, or Medicine) will primarily use quantitative data.
- When presenting quantitative results, authors will typically focus on numbers-based approaches that involve statistics, calculations, and data measurements.
- Example of model evaluation on test data -accuracy calculation., run-time, big O, ROC, ..

RESULTS STRUCTURING

- There are three basic parts for this section:
  - First, begin with an introduction: connecting results with research questions.
  - Second, present your findings in a well structure(thematically or chronologically) with focus on important, interesting, or significant findings combining text and visuals.
  - Third, the results section closing summary of the key findings against the research questions and existing literature.

[Introductory Context] A total of 5365 snakebites were reported to the CPCS from 1 September 1997 through 30 September 2017. All bites were reported from rattlesnakes. ***[Relevant Finding]*** The majority of snakebite reports were called from health care facilities (4607, 85.9%) versus private residences (671, 12.6%), with the distribution of number of cases per county and incidence (number of cases per 1 million residents) shown in Figures 1 and 2. [...] ***[Significant Finding (including a significant test result)]*** Males were significantly more likely than females to be injured (76.6%, 95% CI ¼ 75.6–77.9%, p < .01) [...] ***[Reference to Visual Data]*** Table 1 describes the demographics, snakebites, call sites, severity, and treatments. [...] Among the 3580 (67%) of reports with detailed treatment data, the most common intervention was FAb antivenom (59%). ***[Relevant Finding]*** Controlling for population showed that increased population density equated to greater number of snakebites (Supplementary Appendix Figure 1), however, rural areas where snakes natural habitat is more common had a greater incidence of snakebites per capita (Supplementary Appendix Figure 2). ***[Context to Research Question]*** Season and trend decomposition were used to explore snakebite incidence over the study duration and interrelated with drought, precipitation, and temperature. It is assumed that high temperatures and low precipitation predispose drought, while the opposite is true for the recovery from drought. ***[Explanation of Test]*** All three variables were tested independently to determine which indicator provided the best predictive accuracy of incidence and on what time scale. The incidence of snakebite per million people declined during drought (mean in drought ¼ 15.10, mean outside drought ¼ 18.57, 95% CI difference ¼ 0.12–6.83, p ¼ .04). [...] ***[Significant Findings]*** Statistically significant correlations were found between both lagged precipitation (i.e., precipitation from previous months) and snakebite incidence (maximum rho ¼ –0.52, p .01 at 18 months prior) and drought and incidence (maximum rho ¼ 0.17, p ¼ .01 at 6 months prior). […] ***[Summary of Key Findings]*** Taking each predictive variable in isolation (due to their mutual dependence) and adjusting for seasonal fluctuations, excess snakebite incidence could be calculated directly from the six month prior drought at a rate of –3.8% per þ10% increase in drought, and þ3.9% for each 10% increase in 18 month prior precipitation across the state of California (Figure 8) and counties.

Source: Phillips C., Lipman G., Gugelmann H, Doering K., & Lung D., "Snakebites and climate change in California, 1997-2017," *Clinical Toxicology*, 2019.

RESULTS DOS AND DON'TS

Do take time to establish key findings in connection to your research question.

Do include any negative findings, as failing to do so would hurt your credibility as a scholar.

Do include statistical significance tests if applicable.

Don't present data that is not relevant to your research question(s).

Don't present the same data multiple times, but decide on one format to best convey it.

Don't present raw data that can be summarized or presented visually.

Don't use vague terms or generalizations when presenting your findings, always be specific.

Don't attempt to discuss, interpret or analyze your findings within the results section.

FEW BEST PRACTICES

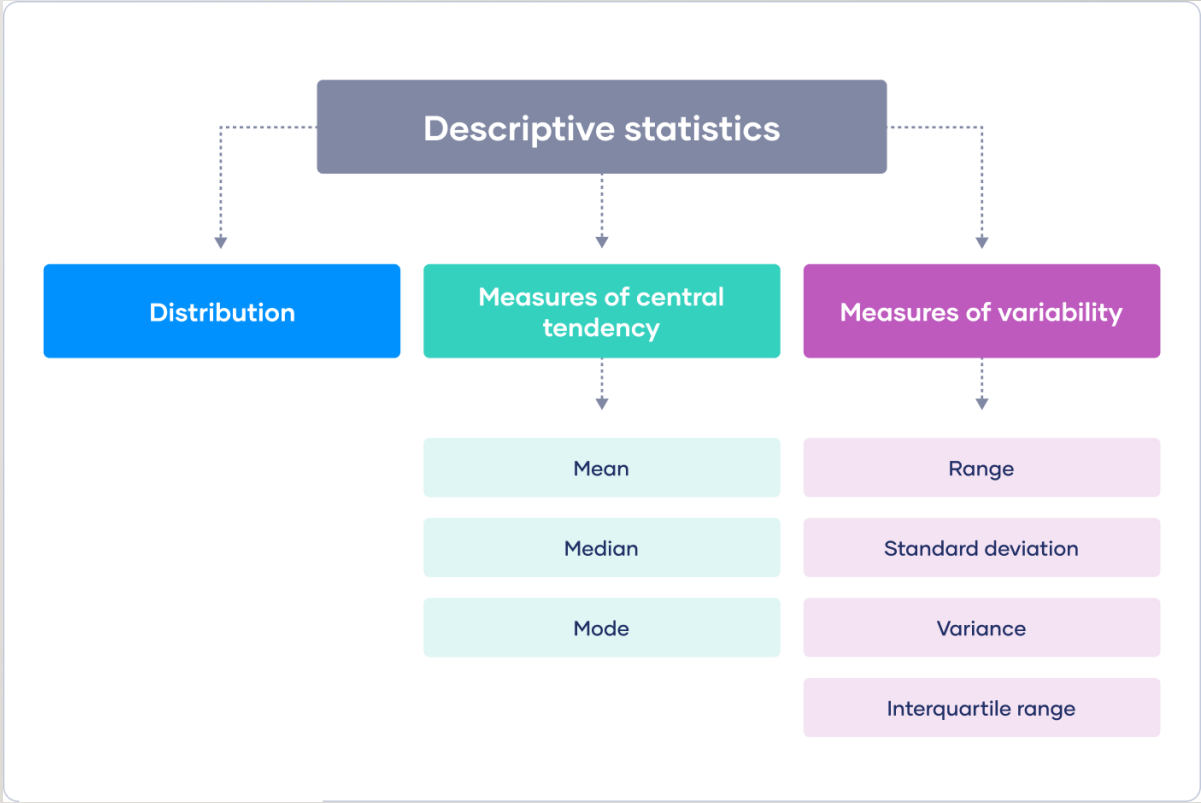
- Your results should always be written in the past tense.
- Only include results that are directly relevant to answering your research questions.
- Avoid speculative or interpretative words like “appears” or “implies.
- Always start out with your broadest results first, and then flow into your more granular (but still relevant) ones.

STATISTICAL ANALYSIS

- Statistical analysis means investigating trends, patterns, and relationships using quantitative data.
- You need to specify your hypotheses and make decisions about your research design, sample size, and sampling procedure.
  - Sample causal research question: does the adaptive control of learning rate enhance training performance of ML models?
- After collecting data from your sample, you can organize and summarize the data using descriptive statistics.
- Then, you can use inferential statistics to formally test hypotheses and make estimates about the population.
- Finally, you can interpret and generalize your findings.

DESCRIPTIVE STATISTICS

- A data set is a collection of responses or observations from a sample or entire population.
- **Descriptive statistics** summarize and organize characteristics of a data set or performance indicator of a system/model.
  - The **distribution** concerns the frequency of each value.
  - The **central tendency** concerns the averages of the values.
  - The **variability** or dispersion concerns how spread out the values are.
- Can be applied in univariate analysis, or to compare two or more, in bivariate and multivariate analysis.



DESCRIPTIVE STATISTICS

- Suppose you are benchmarking a set of classifiers on subsets of test data.

```
for fold, (train_idx, test_idx) in enumerate(kf.split(X_encoded, y)):
    y_train, y_test = y[train_idx], y[test_idx]
    for name, build_fn in models.items():
        model = build_fn()
        model.compile(optimizer='adam', loss='binary_crossentropy' )
        model.fit(X_train, y_train, epochs=5, batch_size=32, verbose=0)
        y_prob = model.predict(X_test).ravel()
        metrics = compute_metrics(y_test, y_prob)
```

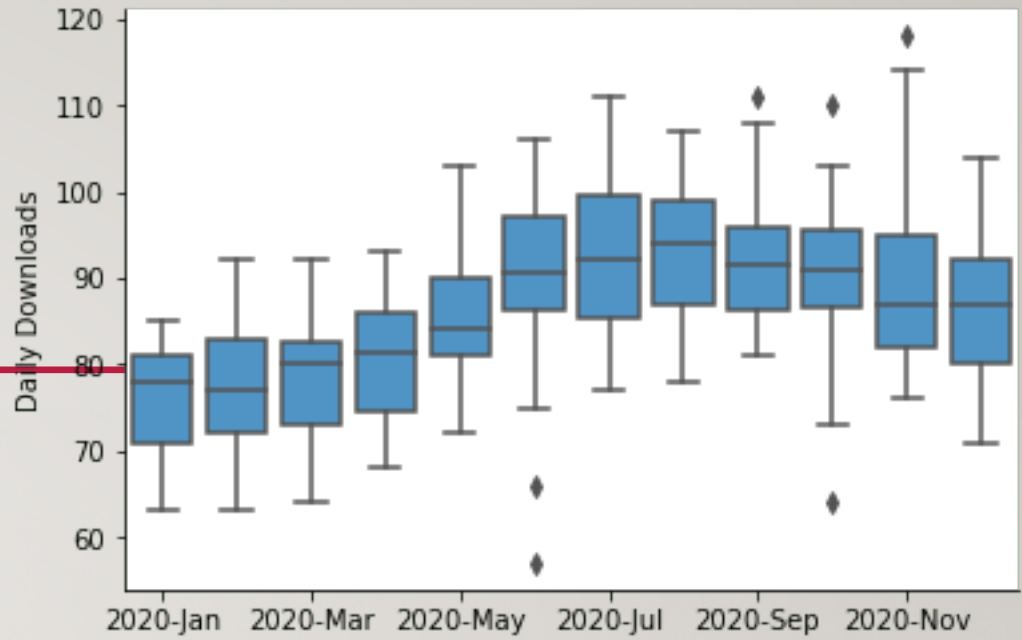
- Now you can use descriptive statistics to find out the averages for each activity (central tendency), and the spread of responses for each activity (variability).
- Then you can measure per model: mean, median, min, max, std, …

APPLIED EXAMPLE

- IMAGINE A PROPOSED MODEL VS. BASELINE MODEL
- YOU ASSESSED ACCURACY OVER 10 TEST FOLDS
- THEN:
  - **Calc. average accuracy over 10 trials**
  - **Standard deviation to prove reliability and may be robustness.**
  - **Best and worst case: help the reader understand the floor and ceil of a criterion.**
- WHAT IF YOU ALSO MEASURED ROC, RUN-TIME, …

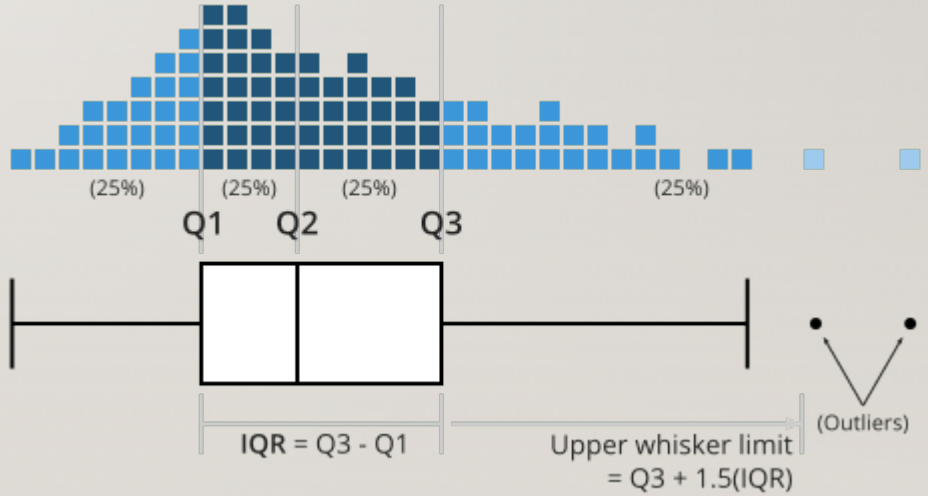
Metric	Baseline Model (CNN)	Proposed Model (ResNet-SCA)	Improvement (Δ)
Mean Accuracy	84.20%	96.50%	+12.3%
Std. Deviation (σ)	3.10%	0.80%	–2.3% (More Stable)
Training Time (per epoch)	45s	58s	+13s
Parameter Count	1.2M	850k	–29% (Lighter)

ALTERNATIVE VISUALIZATION



- A box plot uses boxes and lines to depict the distributions of one or more groups of numeric data.

- Box limits indicate the range of the central 50% of the data, with a central line marking the median value. Lines extend from each box to capture the range of the remaining data, with dots placed past the line edges to indicate outliers.



DESCRIPTIVE STATISTICS

- **Univariate descriptive statistics:**

  Univariate descriptive statistics focus on only one variable at a time.

- **Bivariate descriptive statistics:** you simultaneously study the frequency and variability of two variables to see if they vary together (Scatter plots)

- **Multivariate** analysis is the same as bivariate analysis but with more than two variables.

INFERENTIAL STATISTICS

- **Inferential statistics** help you come to conclusions and make predictions based on your data.
- Inferential statistics use your sample to make reasonable guesses about the larger population.
- I want to prove that ‘my proposed model is significantly better in accuracy over baseline modelA’.

INFERENTIAL STATISTICS

- **Hypothesis testing** is a formal process of statistical analysis using inferential statistics.
- The goal of hypothesis testing is to compare populations between variables using samples.
- Statistical tests can be parametric or non-parametric, with parametric tests more powerful.
- Parametric tests make assumptions that include the following:
  - Sample comes a normal distribution of scores
  - Large sample size
  - Variance of the different sets are similar
- Statistical tests come in three forms: tests of comparison, correlation or regression.

INFERENTIAL STATISTICS: COMPARISON TEST

- **Comparison tests** assess whether there are differences in means, medians or rankings of scores of two or more groups.
- Decide based on your data whether you will go with parametric or non.
- Means can only be found for interval or ratio data, while medians and rankings are more appropriate measures for ordinal data.

Fold	Model A	Model B
1	0.90	0.87
2	0.88	0.86
3	0.91	0.89
4	0.89	0.88
5	0.91	0.999

INFERENTIAL STATISTICS: COMPARISON TEST

- T-tests are used when comparing the means of precisely two groups (e.g., the average heights of men and women).

- Test folds are the same for 2 models ➔ paired t-test {k-fold}
- Both models are tried on different test sets ➔ independent t-test.

	Predictor variable	Outcome variable	Research question example
Paired t-test	• Categorical • 1 predictor	• Quantitative • groups come from the same population	What is the effect of two different test prep programs on the average exam scores for students from the same class?
Independent t-test	• Categorical • 1 predictor	• Quantitative • groups come from different populations	What is the difference in average exam scores for students from two different schools?

INFERENTIAL STATISTICS: COMPARISON TEST –T TEST

- If you only care whether the two populations are different from one another, perform a **two-tailed *t* test**.
- If you want to know whether one population mean is greater than or less than the other, perform a **one-tailed *t* test. [common]**

- *P* values are often interpreted as your risk of rejecting the null hypothesis of your test when the null hypothesis is actually true.
- Always we reject null hypothesis when $p<0.05$

t test example

In your test of whether petal length differs by species:

- Your observations come from two separate populations (separate species), so you perform a two-sample *t* test.
- You don't care about the direction of the difference, only whether there is a difference, so you choose to use a two-tailed *t* test.

INFERENTIAL STATISTICS: COMPARISON TEST

- ANOVA and MANOVA tests are used when comparing the means of more than two groups (e.g., the average heights of children, teenagers, and adults).

ANOVA	• Categorical • 1 or more predictor	• Quantitative • 1 outcome	What is the difference in average pain levels among post-surgical patients given three different painkillers?
MANOVA	• Categorical • 1 or more predictor	• Quantitative • 2 or more outcome	What is the effect of flower species on petal length, petal width, and stem length?

INFERENTIAL STATISTICS: COMPARISON TEST –ANOVA

- ANOVA is the same as t-test while it runs on more than 2 groups.

```
from scipy import stats

group1 = [70, 75, 80]
group2 = [85, 88, 90]
group3 = [60, 65, 70]

F, p = stats.f_oneway(group1, group2, group3)

print("F-value:", F)
print("p-value:", p)
```

INFERENTIAL STATISTICS: COMPARISON TEST

- Non-parametric tests don't make as many assumptions about the data, and are useful when one or more of the common statistical assumptions are violated.
- However, the inferences they make aren't as strong as with parametric tests.
- Commonly when number of samples are so small {small k for k-fold}

	Predictor variable	Outcome variable	Use in place of...
Kruskal–Wallis *H*	• Categorical • 3 or more groups	• Quantitative	ANOVA
ANOSIM	• Categorical • 3 or more groups	• Quantitative • 2 or more outcome variables	MANOVA
Wilcoxon Rank-Sum test	• Categorical • 2 groups	• Quantitative • groups come from different populations	Independent t-test
Wilcoxon Signed-rank test	• Categorical • 2 groups	• Quantitative • groups come from the same population	Paired t-test

INFERENTIAL STATISTICS

- **Correlation tests** determine the extent to which two variables are associated.
- Although Pearson's r is the most statistically powerful test, Spearman's *r* is appropriate for interval and ratio variables when the data doesn't follow a normal distribution.
- The chi square test of independence is the only test that can be used with nominal variables.

Correlation test	Parametric?	Variables
Pearson's *r*	Yes	Interval/ratio variables
Spearman's *r*	No	Ordinal/interval/ratio variables
Chi square test of independence	No	Nominal/ordinal variables

INFERENTIAL STATISTICS

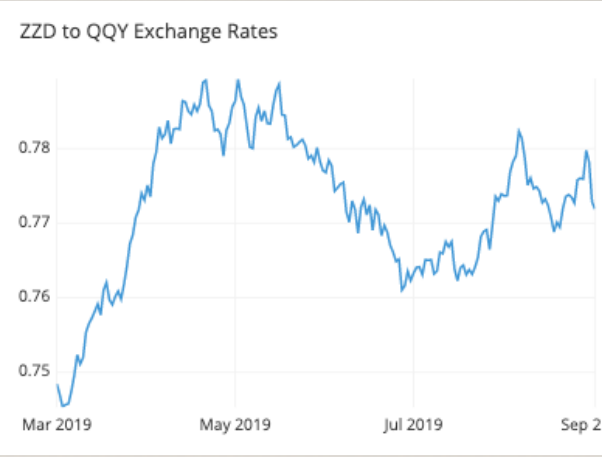
- **Regression tests** demonstrate whether changes in predictor variables cause changes in an outcome variable.
- You can decide which regression test to use based on the number and types of variables you have as predictors and outcomes.
- Most of the commonly used regression tests are parametric. If your data is not normally distributed, you can perform data transformations.
- Data transformations help you make your data normally distributed using mathematical operations, like taking the square root of each value.

INFERENTIAL STATISTICS

Regression test	Predictor	Outcome
Simple linear regression	1 interval/ratio variable	1 interval/ratio variable
Multiple linear regression	2+ interval/ratio variable(s)	1 interval/ratio variable
Logistic regression	1+ any variable(s)	1 binary variable
Nominal regression	1+ any variable(s)	1 nominal variable
Ordinal regression	1+ any variable(s)	1 ordinal variable

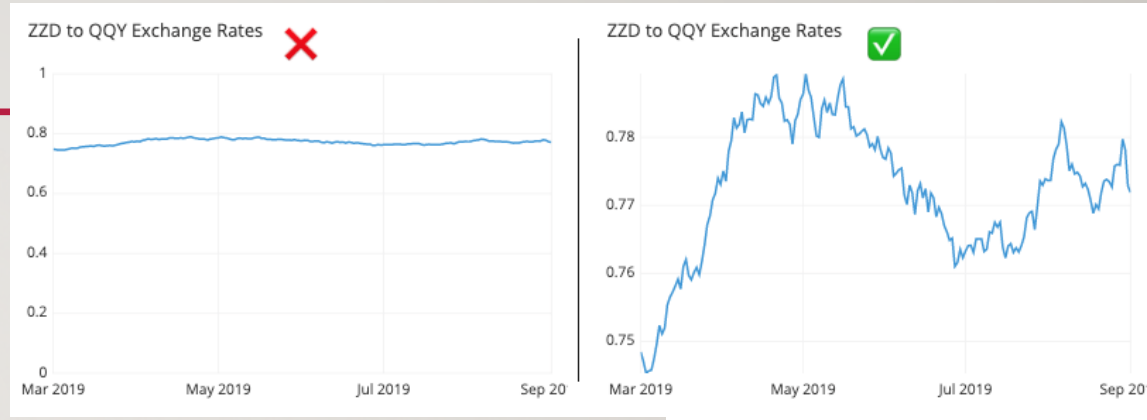
VISUALIZING THE RESULTS

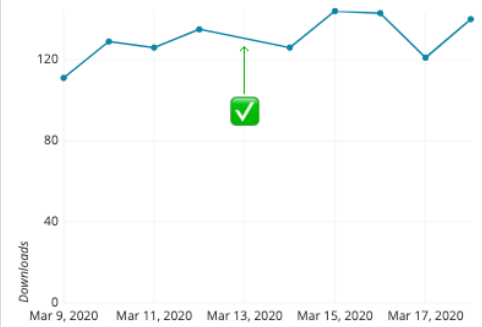
- Line chart: it uses points connected by line segments from left to right to demonstrate changes in value.
- The horizontal axis depicts a continuous progression, often that of time, while the vertical axis reports values for a metric of interest across that progression.



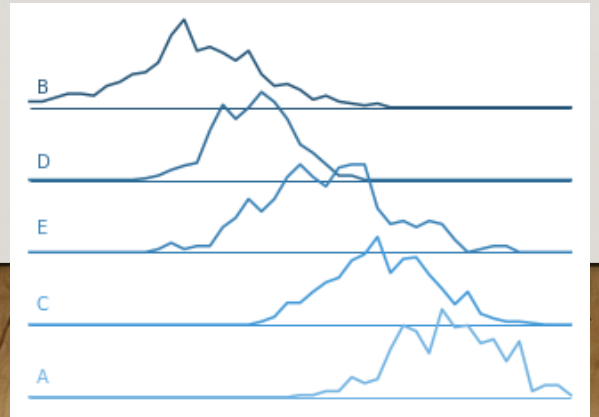
VISUALIZING THE RESULTS

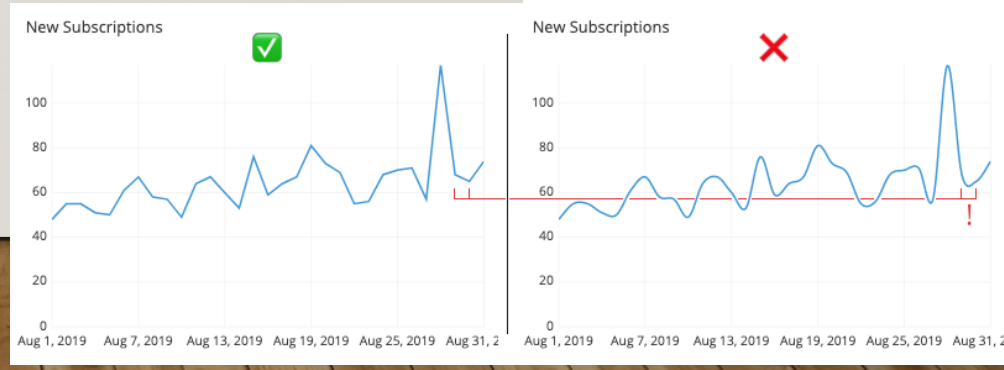
- Common misuses:
  - Strictly using a zero-value baseline
  - Failing to identify uneven gaps between points: When there aren't many points to plot, try showing all of the points and not just the line.
  - Interpolating a curve between points.
  - Using a misleading dual axis: series from different domains





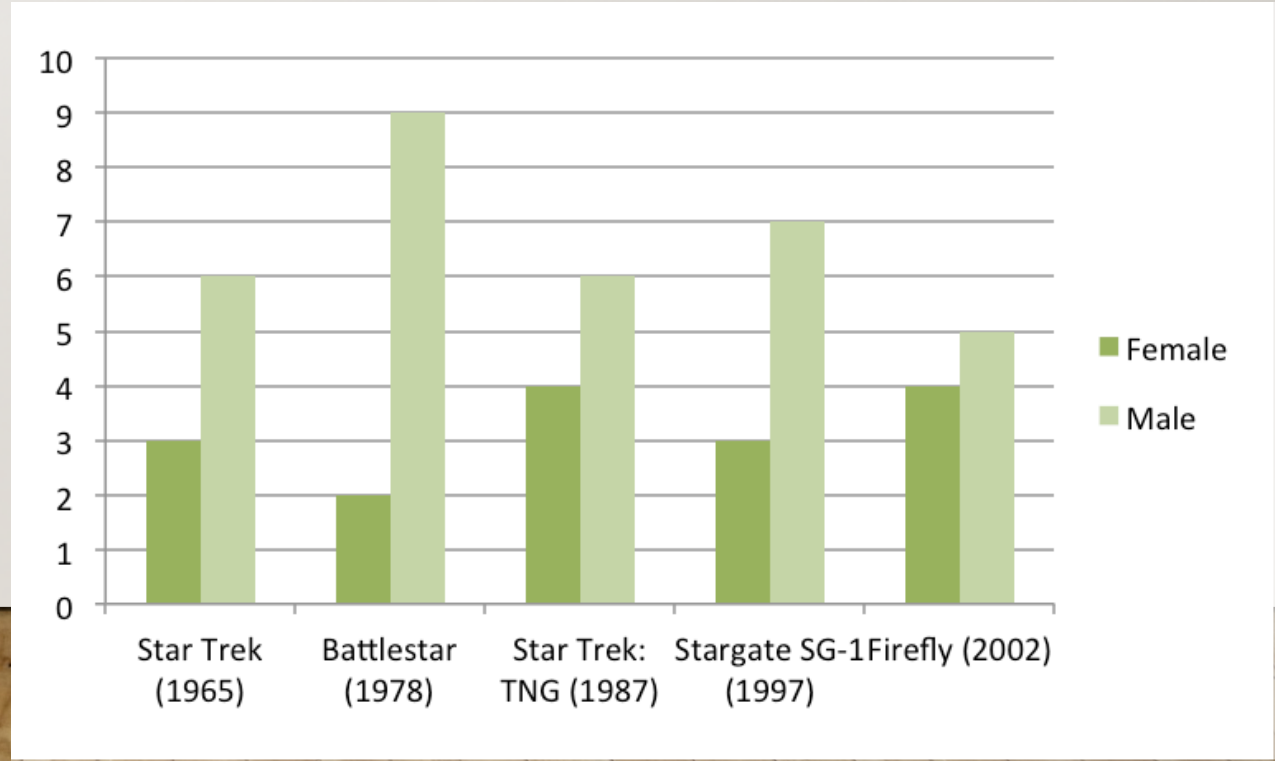
Better to use Ridgeline plot for dual axis plots





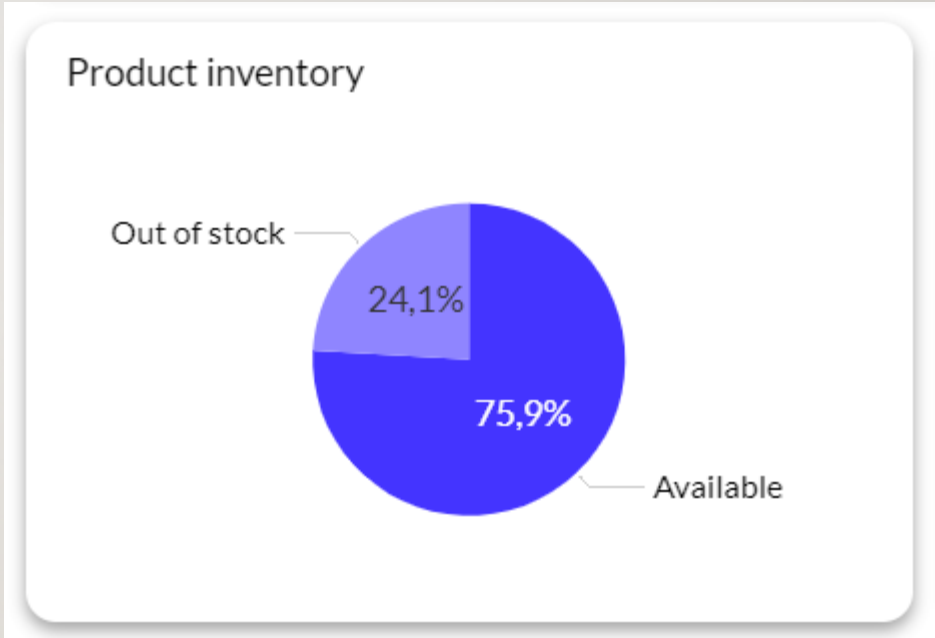
VISUALIZING THE RESULTS

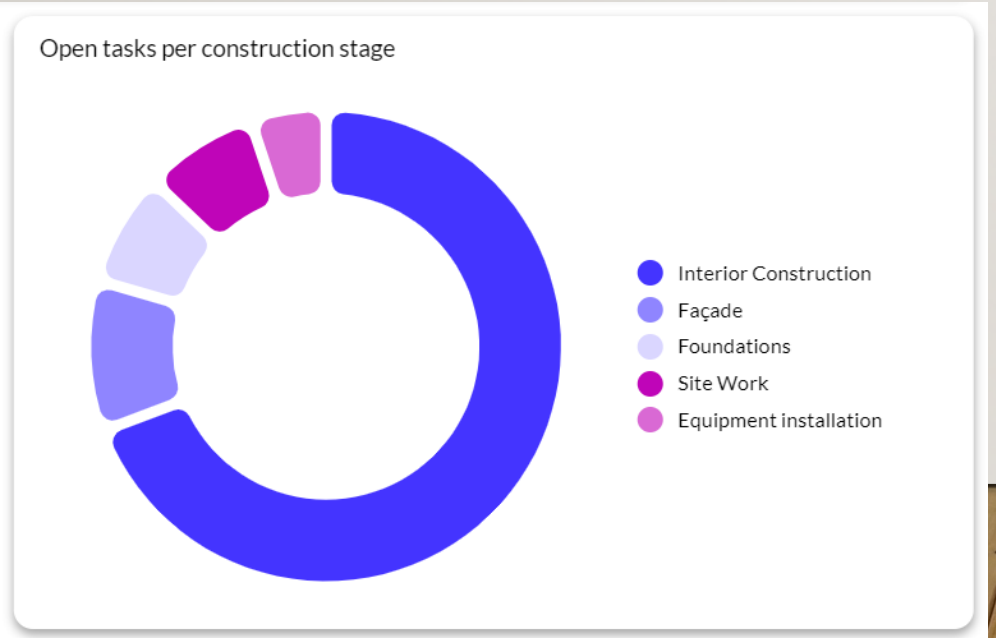
- Bar graphs: they are useful for showing the relationship between independent and dependent variables, where the independent variables are discrete (often nominal) categories.



VISUALIZING THE RESULTS

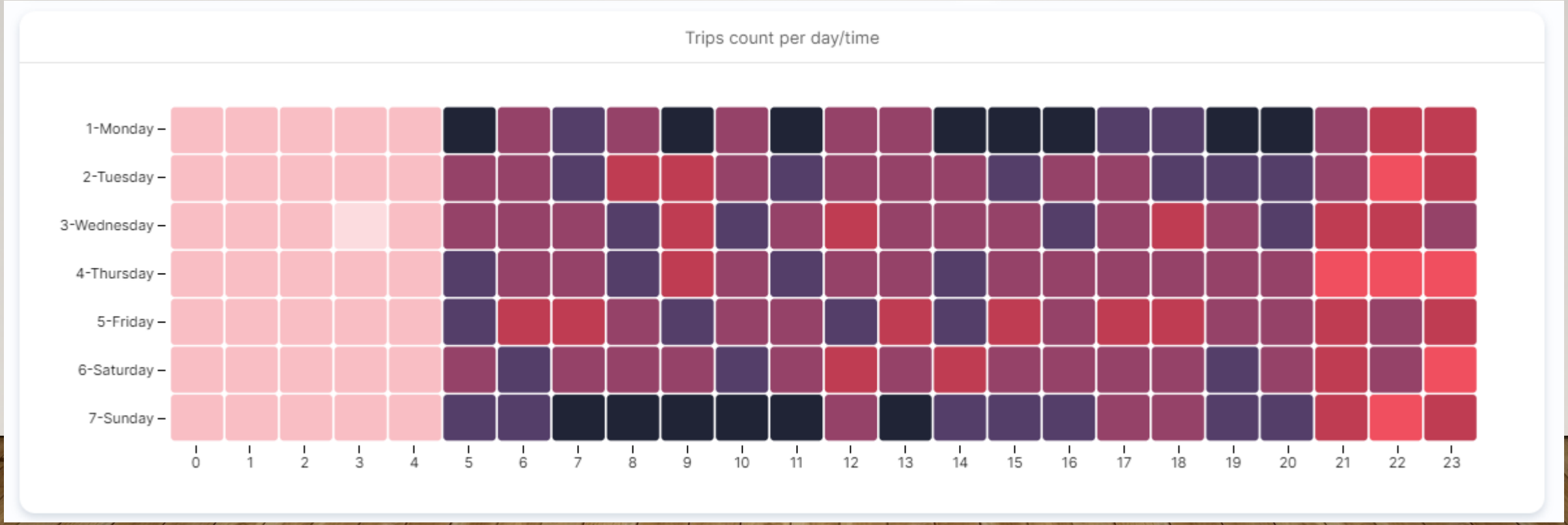
- Pie chart
- Donut chart





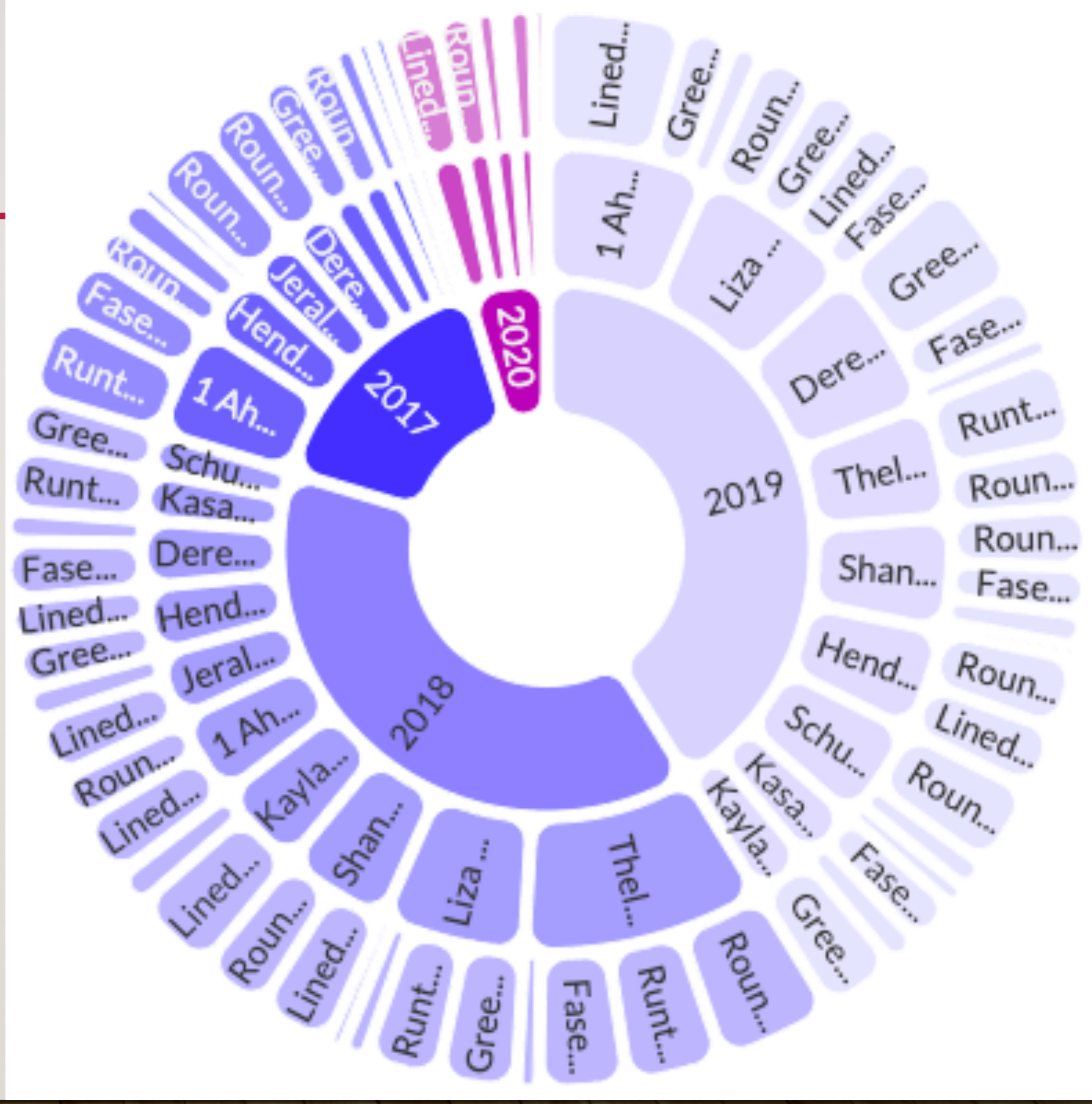
VISUALIZING THE RESULTS

- **Heatmap:** A heatmap or a heat map is a type of chart that shows values for a main variable across two axis variables as a grid with squares in different colors.



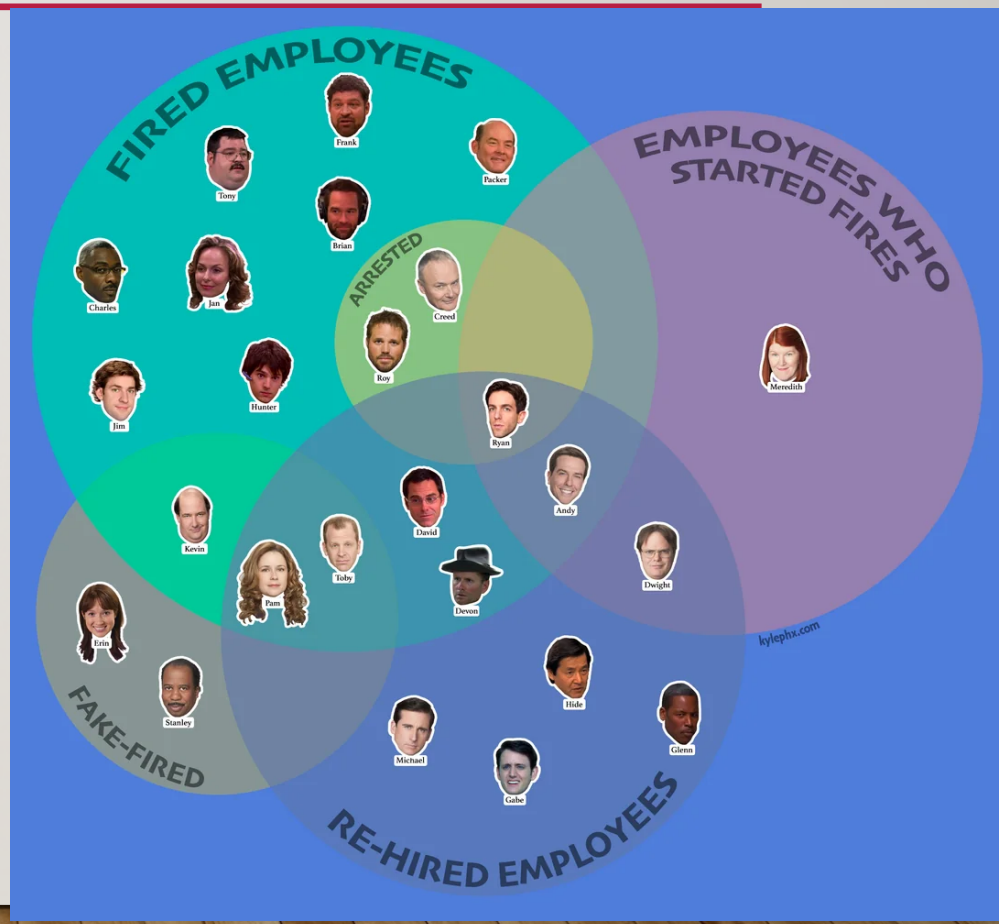
VISUALIZING THE RESULTS

- **Sunburst diagram** shows hierarchical data, and is great for advanced data exploration.
- Instead of using embedded circles, it uses concentric circles.
- For each category you add, you'll add another ring or circle to the visualization.



VISUALIZING THE RESULTS

- **The Venn diagram** is a simple visualization that uses overlapping circles to show relationships between your data.
- Each circle represents one category; the overlap between the circles indicates which characteristics they have in common.



DISCUSSION

- The **discussion section** is where you delve into the meaning, importance, and relevance of your results.
- It should focus on explaining and evaluating what you found, showing how it relates to your literature review and paper or dissertation topic, and making an argument in support of your overall conclusion.
- It should **not** be a second results section.

DISCUSSION

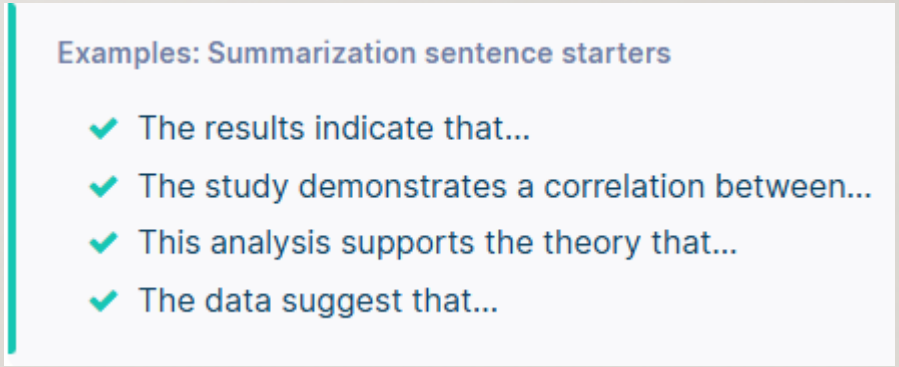
- There are different ways to write this section, but you can focus your writing around these key elements:
  - **Summary:** A brief recap of your key results
  - **Interpretations:** What do your results mean?
  - **Implications:** Why do your results matter?
  - **Limitations:** What can't your results tell us?
  - **Recommendations:** Avenues for further studies or analyses

WHAT NOT TO INCLUDE IN YOUR DISCUSSION SECTION

- **Don’t introduce new results:** You should only discuss the data that you have already reported in your results section.
- **Don’t make inflated claims:** Avoid overinterpretation and speculation that isn’t directly supported by your data.
- **Don’t undermine your research:** The discussion of limitations should aim to strengthen your credibility, not emphasize weaknesses or failures.

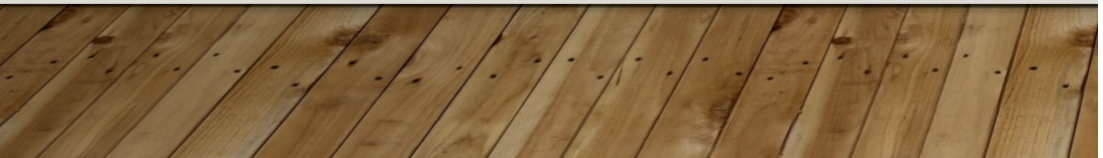
DISCUSSION MAIN STEPS

- **Step 1: Summarize your key findings**
- Start this section by reiterating your research problem and concisely summarizing your major findings.
- This should be no more than one paragraph.



DISCUSSION MAIN STEPS

- **Step 2: Give your interpretations**
- Spell out the significance of results to the reader showing exactly how they answer your research question.
- The form of your interpretations will depend on the type of research, but some typical approaches to interpreting the data include:
  - Identifying correlations, patterns, and relationships among the data
  - Discussing whether the results met your expectations or supported your hypotheses
  - Contextualizing your findings within previous research and theory
  - Explaining unexpected results and evaluating their significance

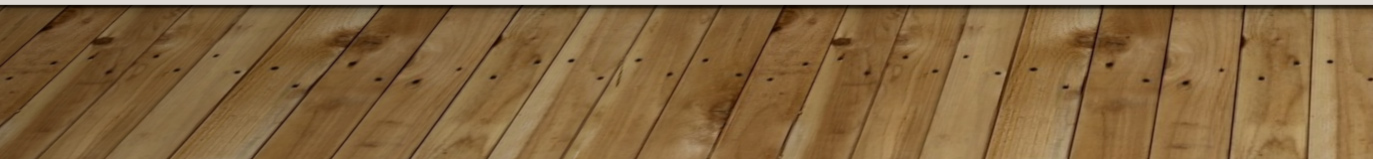


Examples: Interpretation sentence starters

- In line with the hypothesis...
- Contrary to the hypothesized association...
- The results contradict the claims of Smith (2022) that...
- The results might suggest that *x*. However, based on the findings of similar studies, a more plausible explanation is *y*.

DISCUSSION MAIN STEPS

- **Step 3: Discuss the implications**
- Make sure to relate your results back to the scholarly work that you surveyed in the literature review.
- How your findings fit with existing knowledge, what new insights they contribute, and what consequences they have for theory or practice.
- Ask yourself these questions:
  - Do your results support or challenge existing theories? If they support existing theories, what new information do they contribute? If they challenge existing theories, why do you think that is?
  - Are there any practical implications?

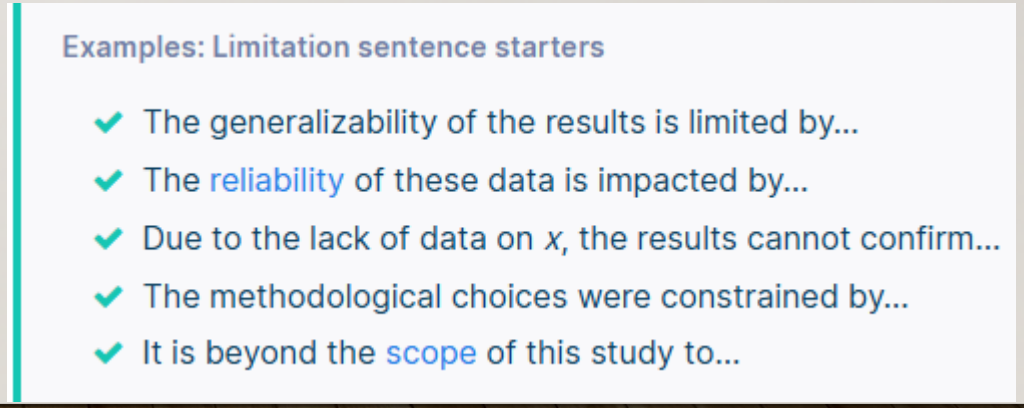


Examples: Implication sentence starters

- These results build on existing evidence of...
- The results do not fit with the theory that...
- The experiment provides a new insight into the relationship between...
- These results should be taken into account when considering how to...
- The data contribute a clearer understanding of...
- While previous research has focused on *x*, these results demonstrate that *y*.

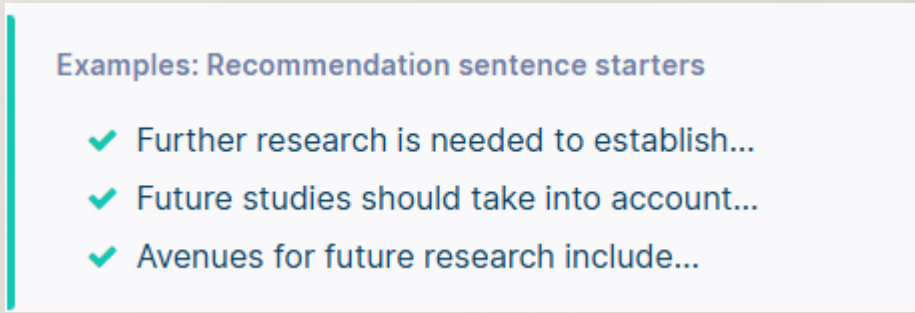
DISCUSSION MAIN STEPS

- **Step 4: Acknowledge the limitations**
- Even the best research has its limitations. Acknowledging these is important to demonstrate your credibility.
- Limitations might be due to your overall research design, specific methodological choices, or unanticipated obstacles that emerged during your research process.



DISCUSSION MAIN STEPS

- **Step 5: Share your recommendations**
- Based on the discussion of your results, you can make recommendations for practical implementation or further research. Sometimes, the recommendations are saved for the conclusion.



REFERENCES

- https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5
- https://www.sjsu.edu/writingcenter/docs/handouts/Results%20Section%20for%20Research%20Papers.pdf
- https://www.scribbr.com/dissertation/results/
- https://www.scribbr.com/dissertation/discussion/
- https://www.scribbr.com/category/statistics/
- https://www.scribbr.com/statistics/statistical-tests/

THANK YOU