Introduction research and data analysis

Felipe Balcazar

NYU

September, 2022

NEW YORK UNIVERSITY

Things to keep in mind

- Come to recitation.
- Book my office hours.
  - Identify problems with your question.
  - Identify problems with your research design.
  - Provide guidance on data, statistics, Stata, R and Python.
- Ask me questions via email.
- Make progress during winter break.
- Don't forget the deadlines!
- Don't wait until the last minute!

Things to keep in mind

- Come to recitation.
- Book my office hours.
  - Identify problems with your question.
  - Identify problems with your research design.
  - Provide guidance on data, statistics, Stata, R and Python.
- Ask me questions via email.
- Make progress during winter break.
- Don't forget the deadlines!
- Don't wait until the last minute!

Things to keep in mind

- Come to recitation.
- Book my office hours.
  - Identify problems with your question.
  - Identify problems with your research design.
  - Provide guidance on data, statistics, Stata, R and Python.
- Ask me questions via email.
- Make progress during winter break.
- Don't forget the deadlines!
- Don't wait until the last minute!

Things to keep in mind

- Come to recitation.
- Book my office hours.
  - Identify problems with your question.
  - Identify problems with your research design.
  - Provide guidance on data, statistics, Stata, R and Python.
- Ask me questions via email.
- Make progress during winter break.
- Don't forget the deadlines!
- Don't wait until the last minute!

Things to keep in mind

- Come to recitation.
- Book my office hours.
  - Identify problems with your question.
  - Identify problems with your research design.
  - Provide guidance on data, statistics, Stata, R and Python.
- Ask me questions via email.
- Make progress during winter break.
- Don't forget the deadlines!
- Don't wait until the last minute!

Things to keep in mind

- Come to recitation.
- Book my office hours.
  - Identify problems with your question.
  - Identify problems with your research design.
  - Provide guidance on data, statistics, Stata, R and Python.
- Ask me questions via email.
- Make progress during winter break.
- Don't forget the deadlines!
- Don't wait until the last minute!

Statistical software

- You will need Stata.
- Purchase your own copy. Obtain Stata SE or MP.
- NYU's computer labs.
- Forums, youtube, ask google.
  - Stata list: `https://www.statalist.org/`
  - UCLA idre: `https://stats.oarc.ucla.edu/`
  - Bobst Library - Data Services: `https://library.nyu.edu/departments/data-services/`

Statistical software

- You will need Stata.
- Purchase your own copy. Obtain Stata SE or MP.
- NYU's computer labs.
- Forums, youtube, ask google.
  - Stata list: `https://www.statalist.org/`
  - UCLA idre: `https://stats.oarc.ucla.edu/`
  - Bobst Library - Data Services: `https://library.nyu.edu/departments/data-services/`

Statistical software

- You will need Stata.
- Purchase your own copy. Obtain Stata SE or MP.
- NYU's computer labs.
- Forums, youtube, ask google.
  - Stata list: `https://www.statalist.org/`
  - UCLA idre: `https://stats.oarc.ucla.edu/`
  - Bobst Library - Data Services: `https://library.nyu.edu/departments/data-services/`

Statistical software

- You will need Stata.
- Purchase your own copy. Obtain Stata SE or MP.
- NYU's computer labs.
- Forums, youtube, ask google.
  - Stata list: `https://www.statalist.org/`
  - UCLA idre: `https://stats.oarc.ucla.edu/`
  - Bobst Library - Data Services: `https://library.nyu.edu/departments/data-services/`

Why Stata?

- Used in academia, govts. and int. organizations.
- Easy to learn.
- Intuitive.
- Many packages - some need to be installed.

Why Stata?

- Used in academia, govts. and int. organizations.
- Easy to learn.
- Intuitive.
- Many packages - some need to be installed.

Why Stata?

- Used in academia, govts. and int. organizations.
- Easy to learn.
- Intuitive.
- Many packages - some need to be installed.

Why Stata?

- Used in academia, govts. and int. organizations.
- Easy to learn.
- Intuitive.
- Many packages - some need to be installed.

We are interested in causal questions

What is the impact of D (your independent variable, the treatment) on Y (the outcome of interest or dependent variable)?

or...

Does D (your independent variable, the treatment) cause Y (the outcome of interest or dependent variable)?

What does this mean?

$$D \longrightarrow Y?$$

Lay out a hypothesis: Define D and Y.

What does this mean?

$$D \longrightarrow Y?$$

Lay out a hypothesis: Define D and Y.

- Formulate the question.
  - Why is it relevant?
  - Why should we care about it?
  - What are we learning from it?

What does this mean?

$$D \longrightarrow Y?$$

Lay out a hypothesis: Define D and Y.

- Formulate the question.
  - Why is it relevant?
  - Why should we care about it?
  - What are we learning from it?

- How are you going to test your hypothesis:
  - What data would you use?
  - What tool will you use to test this hypothesis?
  - How are you going to convince people of your findings?

Developing a Research Question

Choose a topic, define an issue, ask a question, state a hypothesis:

- Topics are general:
  - Civil conflict.
  - Democracy/autocracy.
- Issues are the small components:
  - Climate shocks and unrest in Sub-Saharan Africa.
  - Aid and autocrat survival in developing countries.
- Questions are narrower elements of an issue:
  - Do climate shocks cause conflict?
  - Why does aid help autocrats?
- Propose an educated explanation:
  - Farmers' incomes fall, reducing the cost of rebellion.
  - Fiscal windfalls increase the incentives/means to stay in power.

Developing a Research Question

Choose a topic, define an issue, ask a question, state a hypothesis:

- Topics are general:
  - Civil conflict.
  - Democracy/autocracy.
- Issues are the small components:
  - Climate shocks and unrest in Sub-Saharan Africa.
  - Aid and autocrat survival in developing countries.
- Questions are narrower elements of an issue:
  - Do climate shocks cause conflict?
  - Why does aid help autocrats?
- Propose an educated explanation:
  - Farmers' incomes fall, reducing the cost of rebellion.
  - Fiscal windfalls increase the incentives/means to stay in power.

Developing a Research Question

Choose a topic, define an issue, ask a question, state a hypothesis:

- Topics are general:
  - Civil conflict.
  - Democracy/autocracy.
- Issues are the small components:
  - Climate shocks and unrest in Sub-Saharan Africa.
  - Aid and autocrat survival in developing countries.
- Questions are narrower elements of an issue:
  - Do climate shocks cause conflict?
  - Why does aid help autocrats?
- Propose an educated explanation:
  - Farmers' incomes fall, reducing the cost of rebellion.
  - Fiscal windfalls increase the incentives/means to stay in power.

Developing a Research Question

Choose a topic, define an issue, ask a question, state a hypothesis:

- Topics are general:
  - Civil conflict.
  - Democracy/autocracy.
- Issues are the small components:
  - Climate shocks and unrest in Sub-Saharan Africa.
  - Aid and autocrat survival in developing countries.
- Questions are narrower elements of an issue:
  - Do climate shocks cause conflict?
  - Why does aid help autocrats?
- Propose an educated explanation:
  - Farmers' incomes fall, reducing the cost of rebellion.
  - Fiscal windfalls increase the incentives/means to stay in power.

Sources of inspiration

Puzzles:

- A theoretical/historical/empirical regularity or statement baffling or confusing.

Sources of inspiration

Puzzles:

- A theoretical/historical/empirical regularity or statement baffling or confusing.

Example:

- Puzzle: International aid does not seem to be helping poor countries develop.
- Causal question: Does aid cause corruption in countries with weak institutions?
- Hypothesis: Aid is a fiscal windfall and generates corruption if institutions are weak.

Sources of inspiration

Puzzles:

- A theoretical/historical/empirical regularity or statement baffling or confusing.

Example:

- Puzzle: International aid does not seem to be helping poor countries develop.
- Causal question: Does aid cause corruption in countries with weak institutions?
- Hypothesis: Aid is a fiscal windfall and generates corruption if institutions are weak.

Sources of inspiration

Resolving a debate:

- Contradicting hypotheses or findings, and there is no consensus.

Sources of inspiration

Resolving a debate:

- Contradicting hypotheses or findings, and there is no consensus.

Example:

- Debate: International trade leads to democratization.
- Causal question: Do countries that depend more on tariff revenues democratize?
- Hypothesis: Trade generates tariff revenues, increases the incentives and means to stay in power.

Sources of inspiration

Resolving a debate:

- Contradicting hypotheses or findings, and there is no consensus.

Example:

- Debate: International trade leads to democratization.
- Causal question: Do countries that depend more on tariff revenues democratize?
- Hypothesis: Trade generates tariff revenues, increases the incentives and means to stay in power.

Sources of inspiration

Resolving a debate:

- Contradicting hypotheses or findings, and there is no consensus.

Example:

- Debate: International trade leads to democratization.
- Causal question: Do countries that depend more on tariff revenues democratize?
- Hypothesis: Trade generates tariff revenues, increases the incentives and means to stay in power.

Sources of inspiration

Explain a phenomenon:

- A phenomenon is occurring but we know very little about it.
- Avoid obvious explanations - they are not interesting.

Sources of inspiration

Explain a phenomenon:

- A phenomenon is occurring but we know very little about it.
- Avoid obvious explanations - they are not interesting.

Example:

- Phenomenon: Union political strength has declined in the U.S.
- Causal question: Does automation reduce union strength?
- Hypothesis: Robots increase the opportunity cost of skilled workers to participate in unions.

Sources of inspiration

Explain a phenomenon:

- A phenomenon is occurring but we know very little about it.
- Avoid obvious explanations - they are not interesting.

Example:

- Phenomenon: Union political strength has declined in the U.S.
- Causal question: Does automation reduce union strength?
- Hypothesis: Robots increase the opportunity cost of skilled workers to participate in unions.

Sources of inspiration

Explain a phenomenon:

- A phenomenon is occurring but we know very little about it.
- Avoid obvious explanations - they are not interesting.

Example:

- Phenomenon: Union political strength has declined in the U.S.
- Causal question: Does automation reduce union strength?
- Hypothesis: Robots increase the opportunity cost of skilled workers to participate in unions.

Sources of inspiration

Explain a phenomenon:

- A phenomenon is occurring but we know very little about it.
- Avoid obvious explanations - they are not interesting.

Example:

- Phenomenon: Union political strength has declined in the U.S.
- Causal question: Does automation reduce union strength?
- Hypothesis: Robots increase the opportunity cost of skilled workers to participate in unions.

Theoretical considerations

- Why is the phenomenon taking place?
- What are the incentives of countries/politicians/groups/people?
- Why? When? How?

Theoretical considerations

- Why is the phenomenon taking place?
- What are the incentives of countries/politicians/groups/people?
- Why? When? How?

Theoretical considerations

- Why is the phenomenon taking place?
- What are the incentives of countries/politicians/groups/people?
- Why? When? How?

Theoretical considerations

- Why is the phenomenon taking place?
- What are the incentives of countries/politicians/groups/people?
- Why? When? How?

Example:

- Aid $\Rightarrow$ autocrat survival?

Theoretical considerations

- Why is the phenomenon taking place?
- What are the incentives of countries/politicians/groups/people?
- Why? When? How?

Example:

- Aid $\Rightarrow$ autocrat survival?
- Autocrats are not central planners, they have incentives to maximize their income.
  - Citizens cannot hold autocrats accountable.
  - Cost of being punished is low $\Rightarrow$ extract rents from state and repression.

Theoretical considerations

- Why is the phenomenon taking place?
- What are the incentives of countries/politicians/groups/people?
- Why? When? How?

Example:

- Aid $\Rightarrow$ autocrat survival?
- Autocrats are not central planners, they have incentives to maximize their income.
  - Citizens cannot hold autocrats accountable.
  - Cost of being punished is low $\Rightarrow$ extract rents from state and repression.
- When institutions are weak it is hard to hold leader accountable.

Choosing a good research question

1. Identify a question that's interesting.
   - Ask a causal question.
   - Elevator pitch. You want to communicate and idea/lesson to the reader.
2. Identify a question that you are passionate about.
3. Read the existing literature.
   - We are standing in the shoulders of giants.
   - No need to over do it - but important to cite main contributions.
4. Identify your contribution.
   - What am I saying that is new and interesting?
   - No need to be ground-breaking.

Choosing a good research question

1. Identify a question that's interesting.
   - Ask a causal question.
   - Elevator pitch. You want to communicate and idea/lesson to the reader.
2. Identify a question that you are passionate about.
3. Read the existing literature.
   - We are standing in the shoulders of giants.
   - No need to over do it - but important to cite main contributions.
4. Identify your contribution.
   - What am I saying that is new and interesting?
   - No need to be ground-breaking.

Choosing a good research question

1. Identify a question that's interesting.
   - Ask a causal question.
   - Elevator pitch. You want to communicate and idea/lesson to the reader.
2. Identify a question that you are passionate about.
3. Read the existing literature.
   - We are standing in the shoulders of giants.
   - No need to over do it - but important to cite main contributions.
4. Identify your contribution.
   - What am I saying that is new and interesting?
   - No need to be ground-breaking.

Choosing a good research question

1. Identify a question that's interesting.
   - Ask a causal question.
   - Elevator pitch. You want to communicate and idea/lesson to the reader.
2. Identify a question that you are passionate about.
3. Read the existing literature.
   - We are standing in the shoulders of giants.
   - No need to over do it - but important to cite main contributions.
4. Identify your contribution.
   - What am I saying that is new and interesting?
   - No need to be ground-breaking.

Your are not writing research to win a nobel prize

1. You are probably not going to solve a world problem.
2. Think about something doable for the time frame you have.
3. Even if the contribution is small, that is ok.
4. It is better to be a bit narrow than too broad.
5. It is better to answer a smaller question, and do it well.
6. Research need not be perfect, but good enough.
7. The best papers do their best by contributing to answering a big question, by answering a (much) smaller one.

Your are not writing research to win a nobel prize

1. You are probably not going to solve a world problem.
2. Think about something doable for the time frame you have.
3. Even if the contribution is small, that is ok.
4. It is better to be a bit narrow than too broad.
5. It is better to answer a smaller question, and do it well.
6. Research need not be perfect, but good enough.
7. The best papers do their best by contributing to answering a big question, by answering a (much) smaller one.

Your are not writing research to win a nobel prize

1. You are probably not going to solve a world problem.
2. Think about something doable for the time frame you have.
3. Even if the contribution is small, that is ok.
4. It is better to be a bit narrow than too broad.
5. It is better to answer a smaller question, and do it well.
6. Research need not be perfect, but good enough.
7. The best papers do their best by contributing to answering a big question, by answering a (much) smaller one.

Your are not writing research to win a nobel prize

1. You are probably not going to solve a world problem.
2. Think about something doable for the time frame you have.
3. Even if the contribution is small, that is ok.
4. It is better to be a bit narrow than too broad.
5. It is better to answer a smaller question, and do it well.
6. Research need not be perfect, but good enough.
7. The best papers do their best by contributing to answering a big question, by answering a (much) smaller one.

Your are not writing research to win a nobel prize

1. You are probably not going to solve a world problem.
2. Think about something doable for the time frame you have.
3. Even if the contribution is small, that is ok.
4. It is better to be a bit narrow than too broad.
5. It is better to answer a smaller question, and do it well.
6. Research need not be perfect, but good enough.
7. The best papers do their best by contributing to answering a big question, by answering a (much) smaller one.

Your are not writing research to win a nobel prize

1. You are probably not going to solve a world problem.
2. Think about something doable for the time frame you have.
3. Even if the contribution is small, that is ok.
4. It is better to be a bit narrow than too broad.
5. It is better to answer a smaller question, and do it well.
6. Research need not be perfect, but good enough.
7. The best papers do their best by contributing to answering a big question, by answering a (much) smaller one.

Your are not writing research to win a nobel prize

1. You are probably not going to solve a world problem.
2. Think about something doable for the time frame you have.
3. Even if the contribution is small, that is ok.
4. It is better to be a bit narrow than too broad.
5. It is better to answer a smaller question, and do it well.
6. Research need not be perfect, but good enough.
7. The best papers do their best by contributing to answering a big question, by answering a (much) smaller one.

Relevant literature and reference management

- Google Scholar (Check citations).
- Annual Review of Political Science; Journal of Economic Literature.
- Who are these authors reading and citing?
- Focus on recent takes on the subject (peer reviewed articles and books).
- Keep track because you will cite some of these articles/books.

Relevant literature and reference management

- Google Scholar (Check citations).
- Annual Review of Political Science; Journal of Economic Literature.
- Who are these authors reading and citing?
- Focus on recent takes on the subject (peer reviewed articles and books).
- Keep track because you will cite some of these articles/books.

Relevant literature and reference management

- Google Scholar (Check citations).
- Annual Review of Political Science; Journal of Economic Literature.
- Who are these authors reading and citing?
- Focus on recent takes on the subject (peer reviewed articles and books).
- Keep track because you will cite some of these articles/books.

Relevant literature and reference management

- Google Scholar (Check citations).
- Annual Review of Political Science; Journal of Economic Literature.
- Who are these authors reading and citing?
- Focus on recent takes on the subject (peer reviewed articles and books).
- Keep track because you will cite some of these articles/books.

Relevant literature and reference management

- Google Scholar (Check citations).
- Annual Review of Political Science; Journal of Economic Literature.
- Who are these authors reading and citing?
- Focus on recent takes on the subject (peer reviewed articles and books).
- Keep track because you will cite some of these articles/books.

Relevant literature and reference management

- Google Scholar (Check citations).
- Annual Review of Political Science; Journal of Economic Literature.
- Who are these authors reading and citing?
- Focus on recent takes on the subject (peer reviewed articles and books).
- Keep track because you will cite some of these articles/books.

Tracking and organizing your references:

- Mendeley desktop: `https://www.mendeley.com/`
- JabRef: `https://www.jabref.org/`
- Google scholar cite button.

Important considerations in practice

Variables you need to think about:

- Dependent variable (Y): the outcome.
- Indepedent variable/treatment (D): the cause of the outcome.
  - It can be binary (received a cash transfer or not).
  - The amount of government transfers.
- Confounders: cause both the treatment and the outcome - sources of bias.
  - Observable? Can it be measured?
  - Unobservable - problematic.

$$D\text{: Aid} \longrightarrow Y\text{: Corruption}$$

Important considerations in practice

Variables you need to think about:

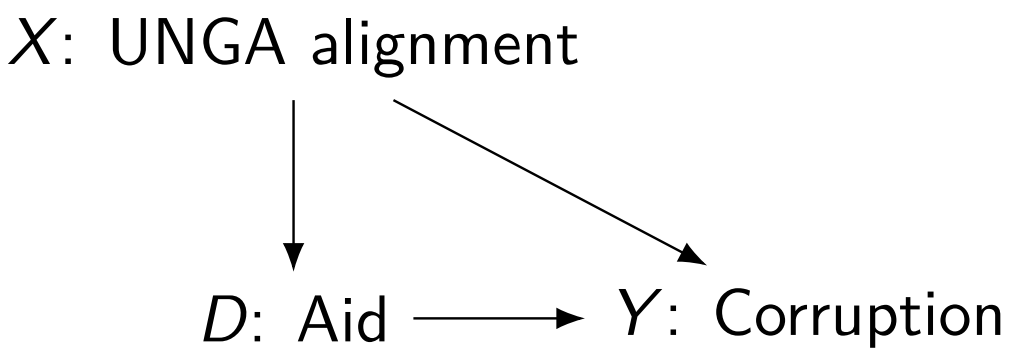
- Dependent variable (Y): the outcome.
- Indepedent variable/treatment (D): the cause of the outcome.
  - It can be binary (received a cash transfer or not).
  - The amount of government transfers.
- Confounders: cause both the treatment and the outcome - sources of bias.
  - Observable? Can it be measured?
  - Unobservable - problematic.

$$D\text{: Aid} \longrightarrow Y\text{: Corruption}$$

Important considerations in practice

Variables you need to think about:

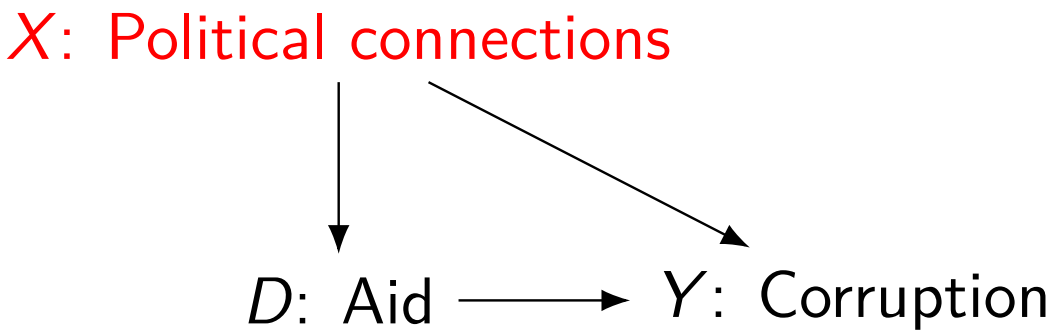
- Dependent variable (Y): the outcome.
- Indepedent variable/treatment (D): the cause of the outcome.
  - It can be binary (received a cash transfer or not).
  - The amount of government transfers.
- Confounders: cause both the treatment and the outcome - sources of bias.
  - Observable? Can it be measured?
  - Unobservable - problematic.



Important considerations in practice

Variables you need to think about:

- Dependent variable (Y): the outcome.
- Indepedent variable/treatment (D): the cause of the outcome.
  - It can be binary (received a cash transfer or not).
  - The amount of government transfers.
- Confounders: cause both the treatment and the outcome - sources of bias.
  - Observable? Can it be measured?
  - Unobservable - problematic.



Very important considerations

Make sure the project is feasible:

- Identify your unit of observation (countries? individuals? counties?).
- Y and D need to vary accross your units of observation.
- If there is a time dimension, Y and D need to vary over time!
- Does your data exist? Can you get it on time?
  - Absence of data breaks a project.
- Can I learn the tool I need in time?
- Do I have the resources buy a dataset or to conduct am experiment/survey, etc.?

Very important considerations

Make sure the project is feasible:

- Identify your unit of observation (countries? individuals? counties?).
- Y and D need to vary accross your units of observation.
- If there is a time dimension, Y and D need to vary over time!
- Does your data exist? Can you get it on time?
  - Absence of data breaks a project.
- Can I learn the tool I need in time?
- Do I have the resources buy a dataset or to conduct am experiment/survey, etc.?

Very important considerations

Make sure the project is feasible:

- Identify your unit of observation (countries? individuals? counties?).
- Y and D need to vary accross your units of observation.
- If there is a time dimension, Y and D need to vary over time!
- Does your data exist? Can you get it on time?
  - Absence of data breaks a project.
- Can I learn the tool I need in time?
- Do I have the resources buy a dataset or to conduct am experiment/survey, etc.?

Very important considerations

Make sure the project is feasible:

- Identify your unit of observation (countries? individuals? counties?).
- Y and D need to vary accross your units of observation.
- If there is a time dimension, Y and D need to vary over time!
- Does your data exist? Can you get it on time?
  - Absence of data breaks a project.
- Can I learn the tool I need in time?
- Do I have the resources buy a dataset or to conduct am experiment/survey, etc.?

Very important considerations

Make sure the project is feasible:

- Identify your unit of observation (countries? individuals? counties?).
- Y and D need to vary accross your units of observation.
- If there is a time dimension, Y and D need to vary over time!
- Does your data exist? Can you get it on time?
  - Absence of data breaks a project.
- Can I learn the tool I need in time?
- Do I have the resources buy a dataset or to conduct am experiment/survey, etc.?

Very important considerations

Make sure the project is feasible:

- Identify your unit of observation (countries? individuals? counties?).
- Y and D need to vary accross your units of observation.
- If there is a time dimension, Y and D need to vary over time!
- Does your data exist? Can you get it on time?
  - Absence of data breaks a project.
- Can I learn the tool I need in time?
- Do I have the resources buy a dataset or to conduct am experiment/survey, etc.?

Off-the shelf datasets

Examples:

- Polity, Variaties of Democracy, Freedom in the World, Correlates of War.
- World Values Survey, Afro/Latino/Euro/Asia/Arab-Barometer, European Social Survey.
- World Penn Tables, IMF Data streams.
- WB World Development indicators.
- IPUMS: `https://www.ipums.org/`
- `https://niehaus.princeton.edu/data`.

Off-the shelf datasets

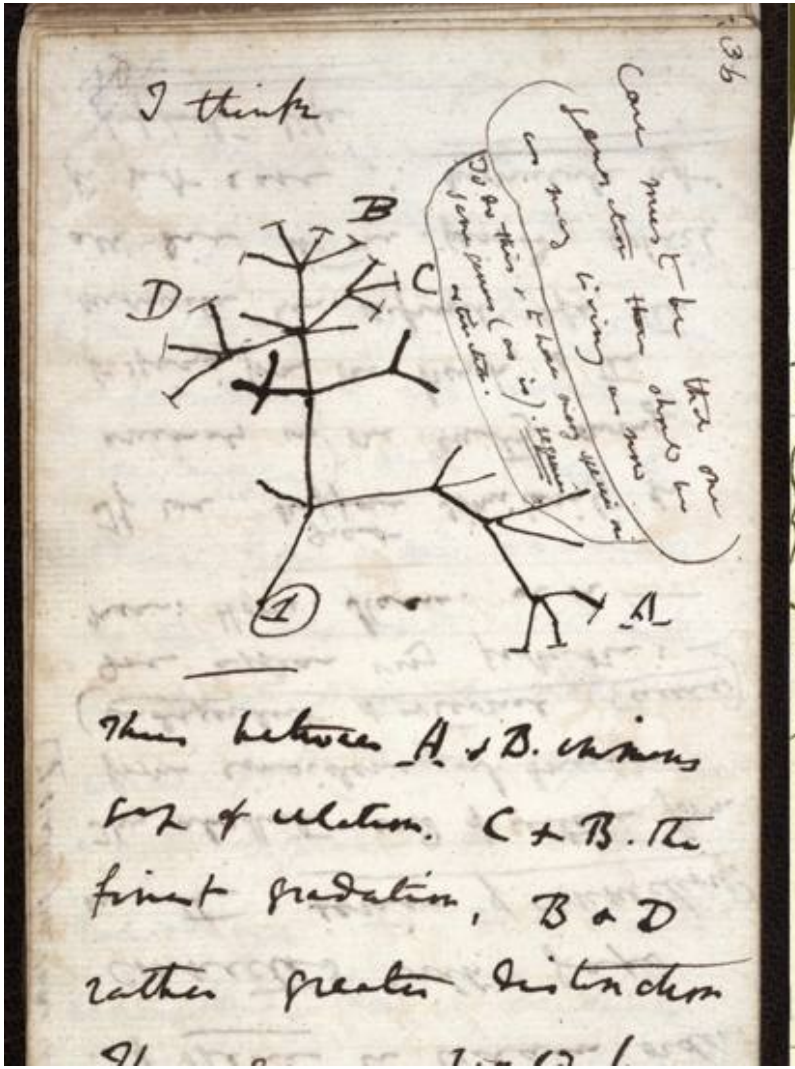
Examples:

- Polity, Variaties of Democracy, Freedom in the World, Correlates of War.
- World Values Survey, Afro/Latino/Euro/Asia/Arab-Barometer, European Social Survey.
- World Penn Tables, IMF Data streams.
- WB World Development indicators.
- IPUMS: `https://www.ipums.org/`
- `https://niehaus.princeton.edu/data`.

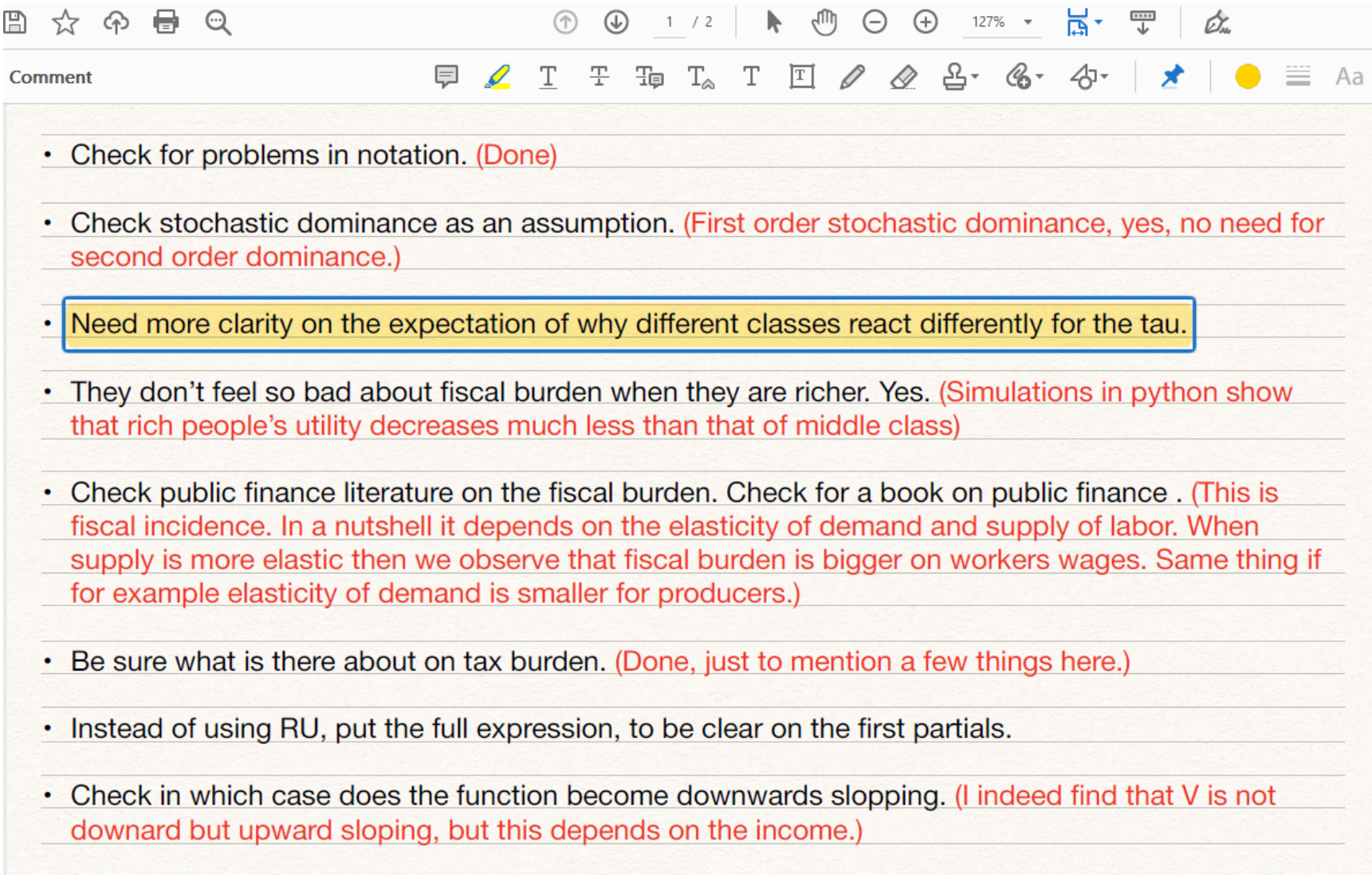
Before using any data:

- What is the source of this data? Is it likely to be accurate? What kinds of biases might there be?
- Read the manuals and codebooks first and check where and how the data previously was used.

A research notebook will save you headaches



A research notebook will save you headaches



Stata

- Open Stata.
- Layout.
- Open a data set.
- Create a variable.
- Check correlation.

Stata

- Open Stata.
- Layout.
- Open a data set.
- Create a variable.
- Check correlation.

Stata

- Open Stata.
- Layout.
- Open a data set.
- Create a variable.
- Check correlation.

Stata

- Open Stata.
- Layout.
- Open a data set.
- Create a variable.
- Check correlation.

Stata

- Open Stata.
- Layout.
- Open a data set.
- Create a variable.
- Check correlation.

- Can you elaborate another question with this data?
- What is the indepdent variable?
- What is the unit of analysis?
- What is the unit of observation?

- Can you elaborate another question with this data?
- What is the indepdent variable?
- What is the unit of analysis?
- What is the unit of observation?

- Can you elaborate another question with this data?
- What is the indepdent variable?
- What is the unit of analysis?
- What is the unit of observation?

- Can you elaborate another question with this data?
- What is the indepdent variable?
- What is the unit of analysis?
- What is the unit of observation?

- Document preparation system, designed for production of scientific and technical docs
- What-You-See-Is-What-You-Get style of document editing (Word etc.) vs. LaTeX
- LaTeXseparates presentation from content $\rightarrow$ LaTeXeditors save us!

Overleaf

1. www.overleaf.com
2. Below "Get started" now, click "Register"
3. To start a new project from scratch, in the main page click the "New Project" $\rightarrow$ "Blank Project"
4. Enter the name $\rightarrow$ "Create"
5. Edit .tex file now & to view the changes click Recompile

Goal: First paragraph of your paper in LaTex!

Overleaf website + excel2latex