

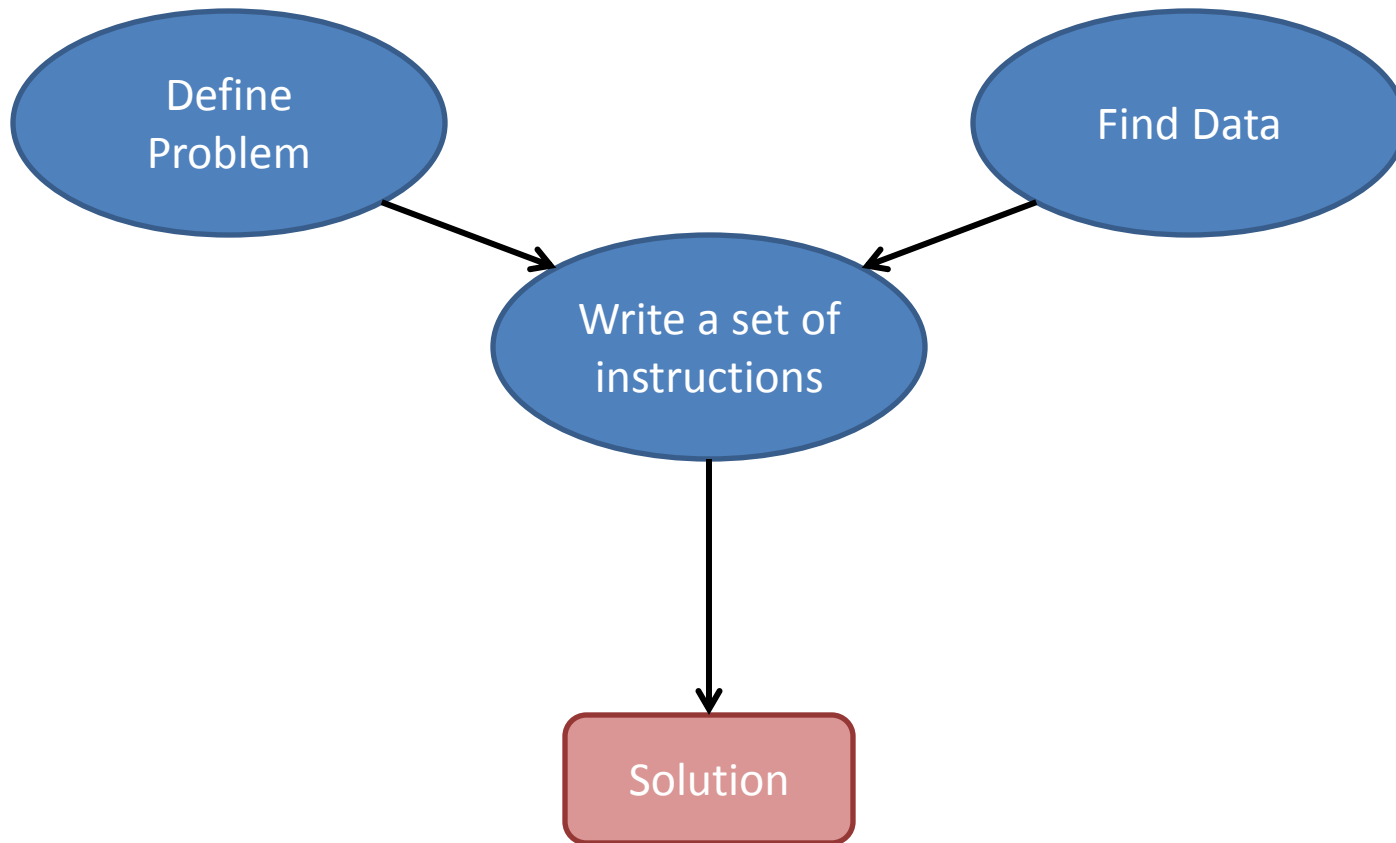
Final Project!

April 16, 2013

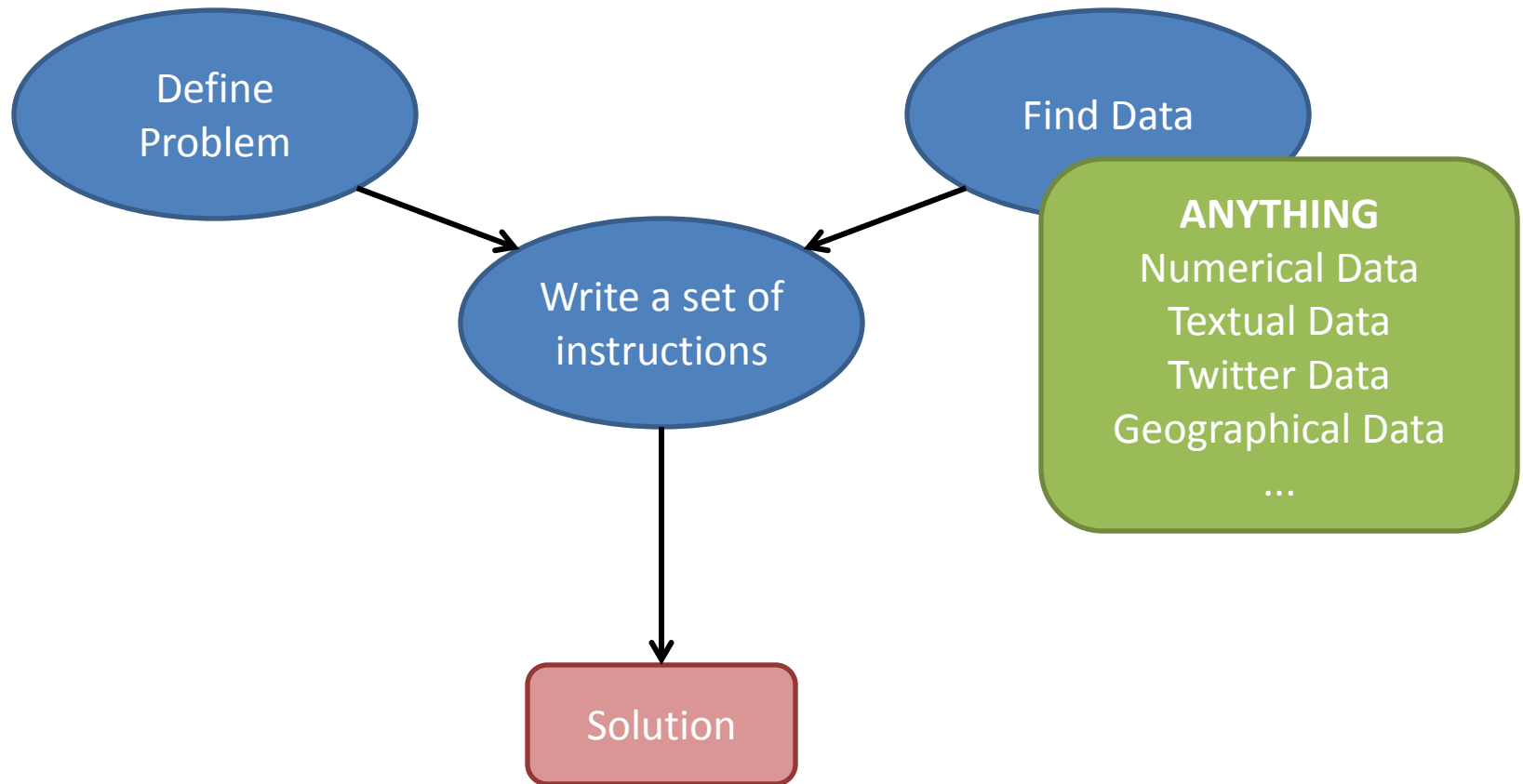
Grading Breakdown

| Component | % of Overall Grade |
|-----------------------------------|--------------------|
| Classroom Participation | 10% |
| Homeworks | 20% |
| Unit Project #1: Voting Patterns | 20% |
| Unit Project #2: Textual Analysis | 20% |
| Final Project | 30% |

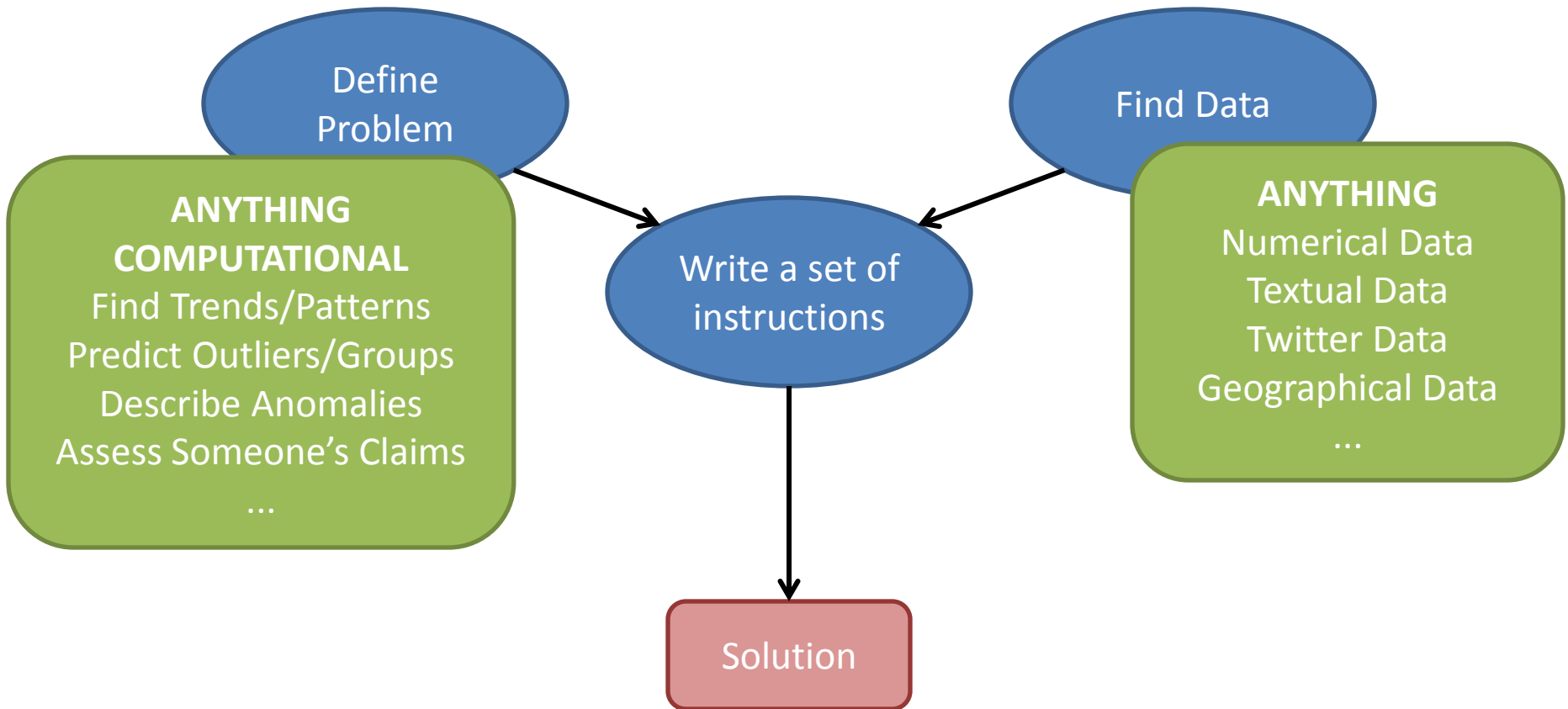
Final Project



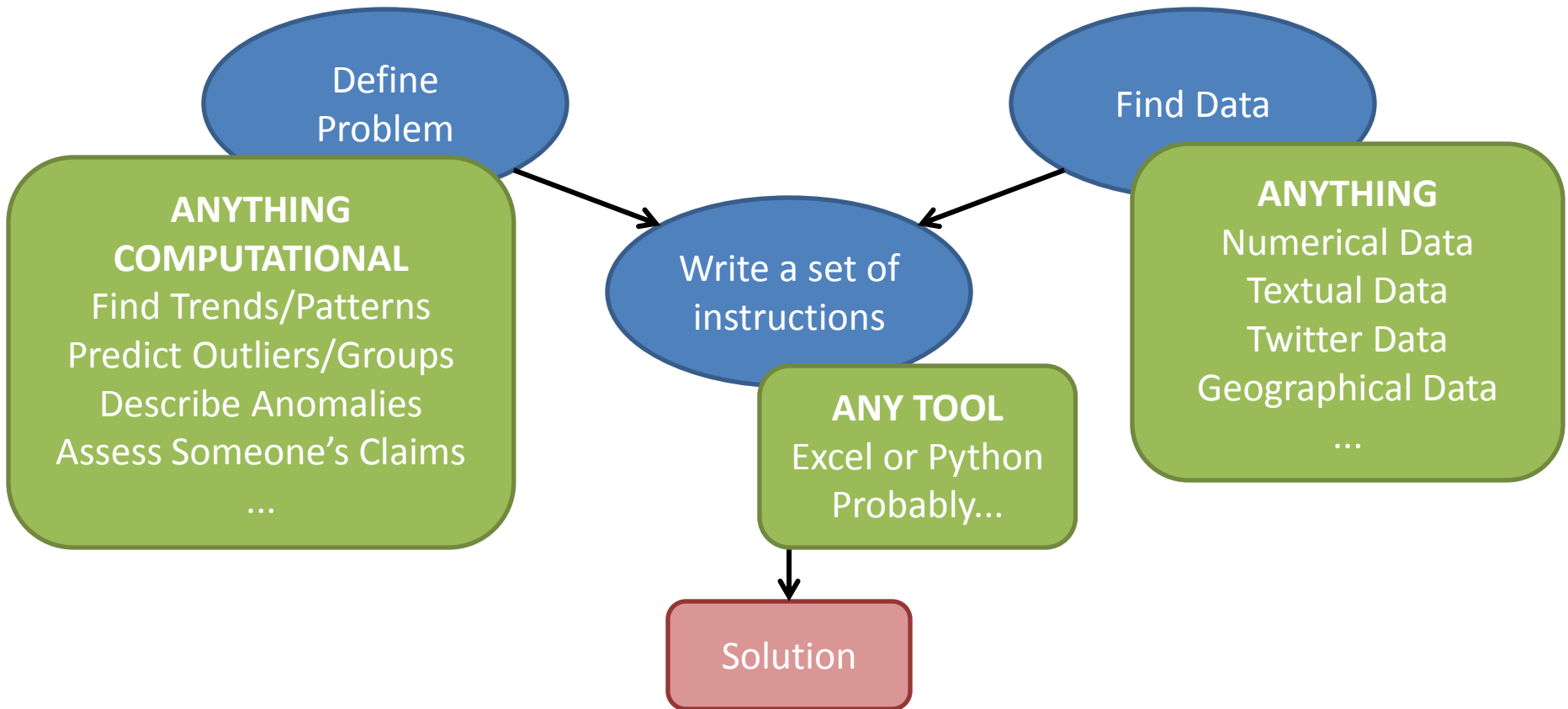
Final Project



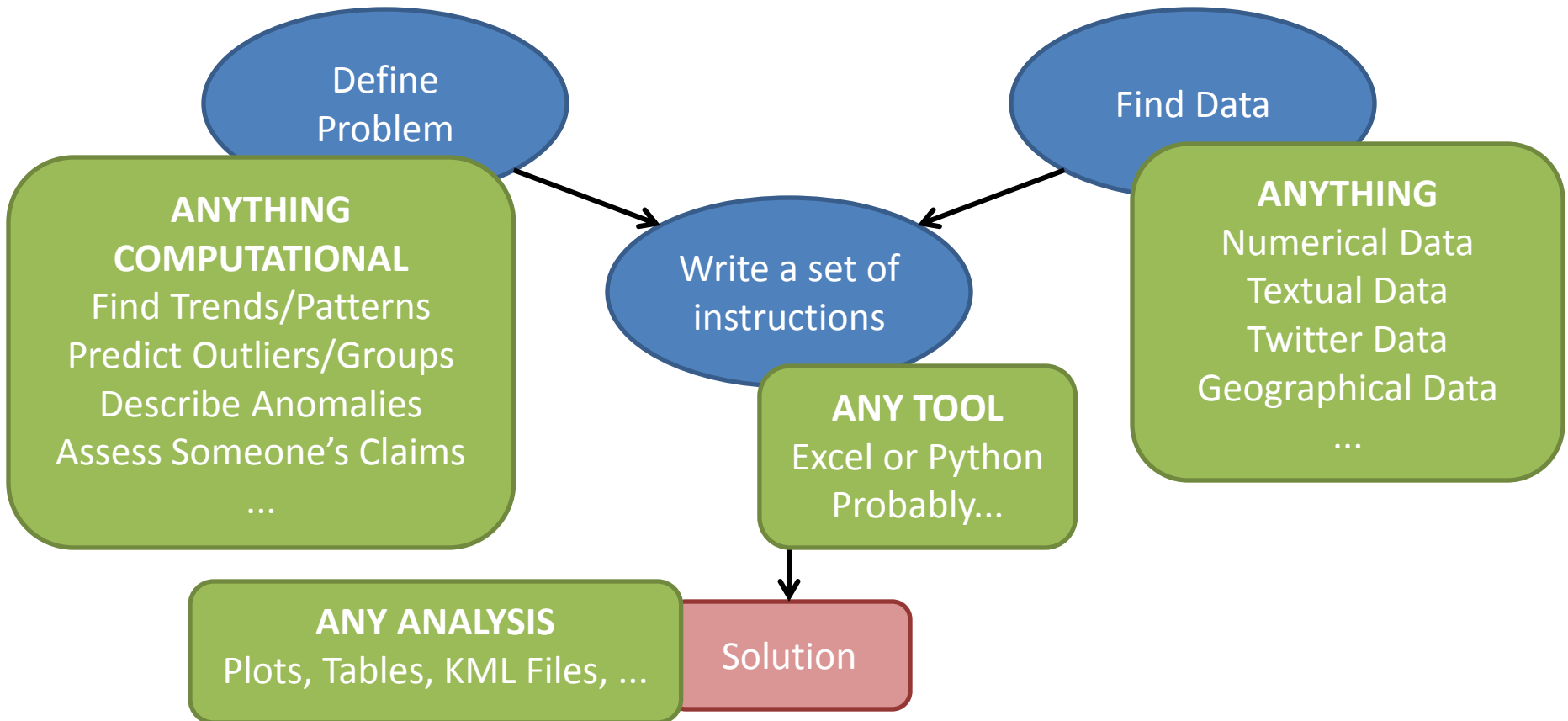
Final Project



Final Project

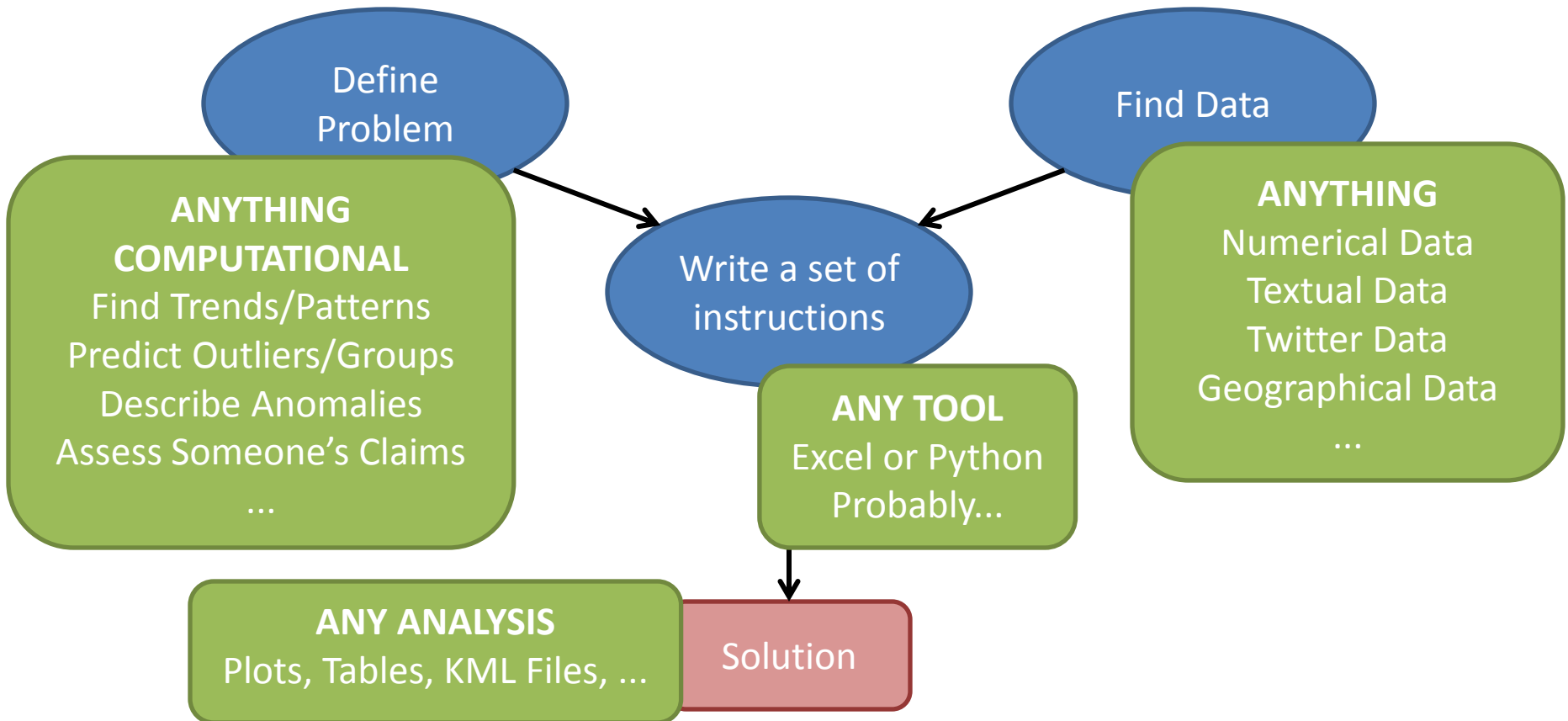


Final Project



Final Project

Only Constant: you will use computers
to process data automatically



Today

- What do we wish we had known before?
- Example Final Projects
- Visitor!
- General Project Description
 - Scope
 - Timeline
- Let's Get Started

What do we wish we
had known before?

Today

- What do we wish we had known before?
- Example Final Projects
- Visitor!
- General Project Description
 - Scope
 - Timeline
- Let's Get Started

Example 1: Where should I Live?

I want to buy a house somewhere near my hometown. Where should I buy?

Kasey Hass's Project

<https://sites.google.com/a/brown.edu/house-hunting-mecklenburg-county/home>

In Each Zip Code:

Rent price: Red dollar sign

Vacancies: For sale sign

Income: Green dollar sign

Marriage rate: Wedding rings



Example 1: Where should I Live?

Kasey Hass's Project

<https://sites.google.com/a/brown.edu/house-hunting-mecklenburg-county/home>

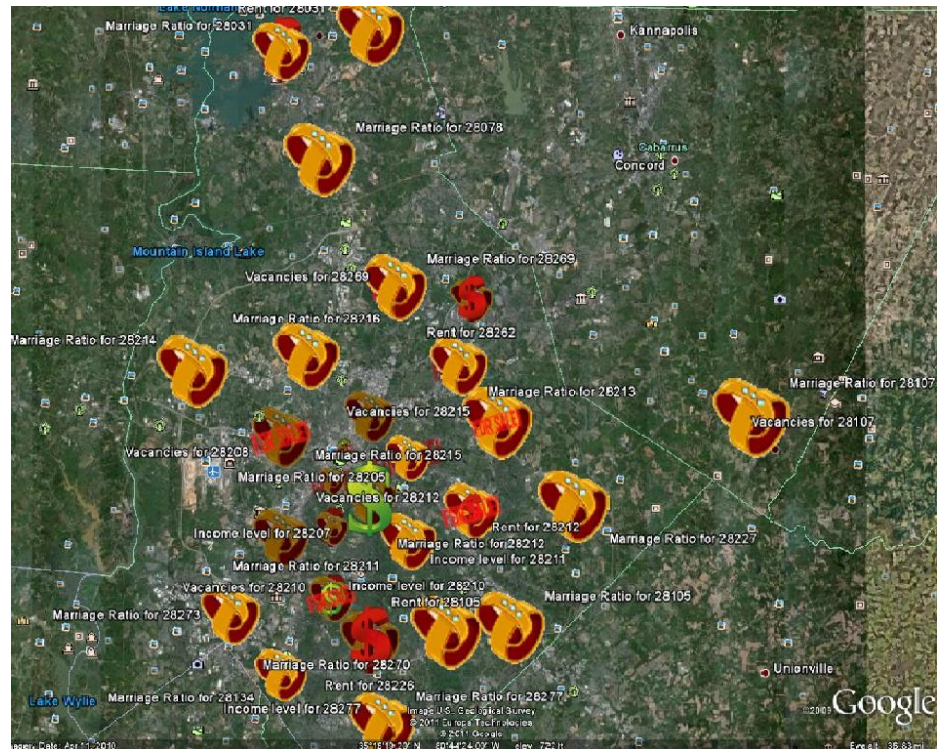
In Each Zip Code:

Rent price: Red dollar sign

Vacancies: For sale sign

Income: Green dollar sign

Marriage rate: Wedding rings



Example 1: Where should I Live?

Kasey Hass's Project

<https://sites.google.com/a/brown.edu/house-hunting-mecklenburg-county/home>

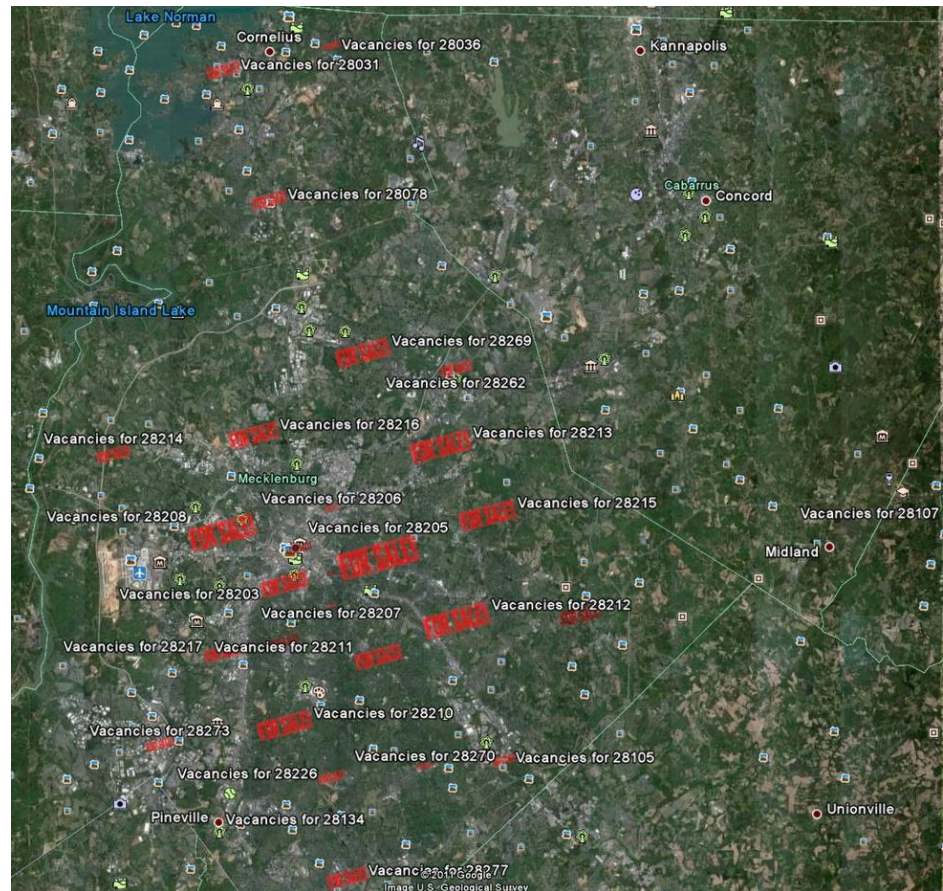
In Each Zip Code:

Rent price: Red dollar sign

Vacancies: For sale sign

Income: Green dollar sign

Marriage rate: Wedding rings



Example 2: Athletes & Geography

Malcolm Gladwell's *Outliers*

- Professional Canadian hockey players

Example 2: Athletes & Geography

Malcolm Gladwell's *Outliers*

- Professional Canadian hockey players
 - Mostly born in January.
 - Few are born in December.
 - Junior league's cutoff is Dec. 31st

Example 2: Athletes & Geography

Malcolm Gladwell's *Outliers*

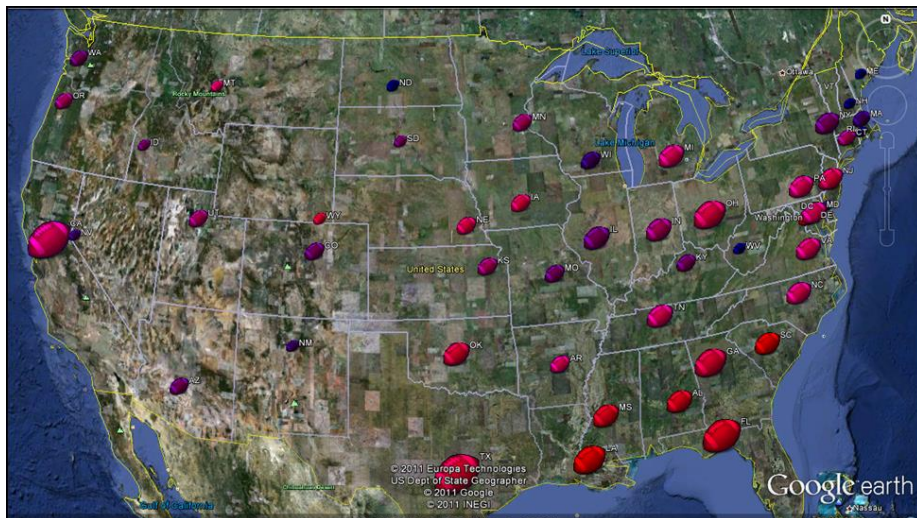
- Professional Canadian hockey players
 - Mostly born in January.
 - Few are born in December.
 - Junior league's cutoff is Dec. 31st
- Any association between professional players and *where* they were born?

Example 2: Athletes & Geography

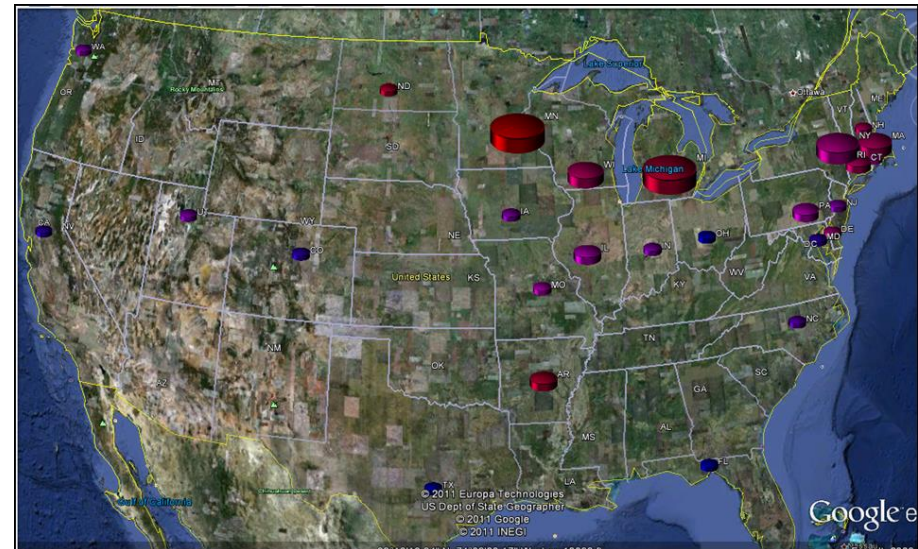
Daniel Newmark's Project

<https://sites.google.com/a/brown.edu/daniel-s-data/project-4--distribution-of-athletes>

Football (NFL)



Hockey (NHL)



Example 3: Growth of Top Digital Media Companies



How does the Growth Rate Compare in the Past 10 Years?

Example 3: Growth of Top Digital Media Companies

Nicholas Talbott's Project

<https://sites.google.com/site/growthofdigitalmediacompanies/home>

One KML File for Each Year
Logo Size shows Growth



YouTube Video:

http://www.youtube.com/watch?v=QDtDF6_mrkc

Visitor!

Gideon Goldin, PhD student in CLPS

Today

- What do we wish we had known before?
- Example Final Projects
- Visitor!
- General Project Description
 - Scope
 - Timeline
- Let's Get Started

Project Description

Reminders

This is an **independent*** assignment *with one exception* (hence the asterisk). Outside of class time, you may discuss this assignment with course staff only. **In class, you may discuss your assignment with anyone** (the instructor, the TAs, or your classmates).

Project Description

(Independent*) Pose a computational question based on numerical data, textual data, or a combination of both. Refer to Projects 1 and 2 for examples of data sources. For your project you must present a testable hypothesis, carry out the required analyses, report your findings in a clear and understandable way, and discuss your results. To present your results, you may use any format(s) you want (a table, a graph, a Google Earth screenshot, etc.). You must make relevant files available on the website (e.g. the KML file for Google Earth).

Project Proposal (Due Thursday, April 23 at 2:25pm)

Project Description

Write a concise (two-page) description of the project you would like to execute. This description should include the following parts:

1. **Background:** put your project idea in context.
2. **Claim:** the specific hypothesis you plan to test (which is a statement, not a question).
3. **Data:** a short description of your data source.
4. **Programming Elements:** a few sentences describing the steps of your analysis and what programs you will use for each (e.g. Excel or Python).
5. **Milestones:** You have three weeks to complete this project. By the first week, you will have this proposal written and some start on the programming/data formatting. What parts do you expect to complete in week 2? What parts do you expect to complete in week 3?
6. **Potential Roadblocks:** a list of potential obstacles.
7. **Project Modifications:** address the following two scenarios:
 - **Backup Plan:** suppose your project is much harder than you anticipate. What parts of the project would you change to still get somewhat interesting results?
 - **Increasing Degree of Difficulty:** suppose your project is much easier than you anticipate. What ways would you extend the project?
8. **Your Goal:** In an ideal world, what will the outcome of this project be? A set of plots displaying a trend? An interactive program? A KML file that users will be able to navigate?

Project (Due Tuesday, May 7 at 2:25pm)

Carry out the project you proposed. It's OK if the project changes – that's why it was a proposal.

Code

Refer to previous project descriptions and rubrics for details about Excel and Python commenting / error handling / etc.

1. You must have a small test dataset (or test Excel sheet) showing that your calculations are correct.
2. You must comment all code (both in Excel and Python).
3. For Excel files, the first sheet must describe all other sheets. Input cells should be highlighted and instructions for use should be clear.
4. For Python files, comments at the top of the program should explain what the program does. Comments throughout the code should make your program easy to follow.

Website

You will also create a website that presents your analysis and results. The site may be accessible only by people with a Brown email address or public. The site should contain the following things:

1. **Project description** and **hypothesis**.
2. **Concise explanation** of your methods.
3. Your **results**, presented in a clear and informative manner.
4. **Discussion of the trends** you see in your analysis. You should point out expected *and* unexpected results.
5. **Reflection** of the project. What went well? What didn't?
6. **Python/Excel/Data files** available for download.

Project (Due Tues, May 7)

Handin

Create a zip file named `YourName_FinalProject.zip`. It should contain a folder that includes the following:

1. All files you used in your project, including Python files, Excel files, data files, and test files.
2. (Optional) A text document named **Supplement.txt** with additional information (tests, analysis, etc) that did not make it on the website.
3. A text file named **README** that contains (1) the url of your web page and (2) a list of all files contained in the zip folder with a short description of what they are.

Hand in the zip file to `cs0931handin@cs.brown.edu`.

Meetings With the Course Staff (no handin)

You must meet with the course staff twice before you hand in your project.

You can go to TA hours or meet during class time. These meetings are intended to help you with your project, and can be brief if you have no unforeseen obstacles.

1. **During week 2** (after you hand in your proposal), you should meet with the course staff **at least once** to discuss your project.
2. **During week 3**, you should meet with the course staff **at least once** to discuss your project.

Interactive Grading (Tuesday, May 7 thru Exam Period)

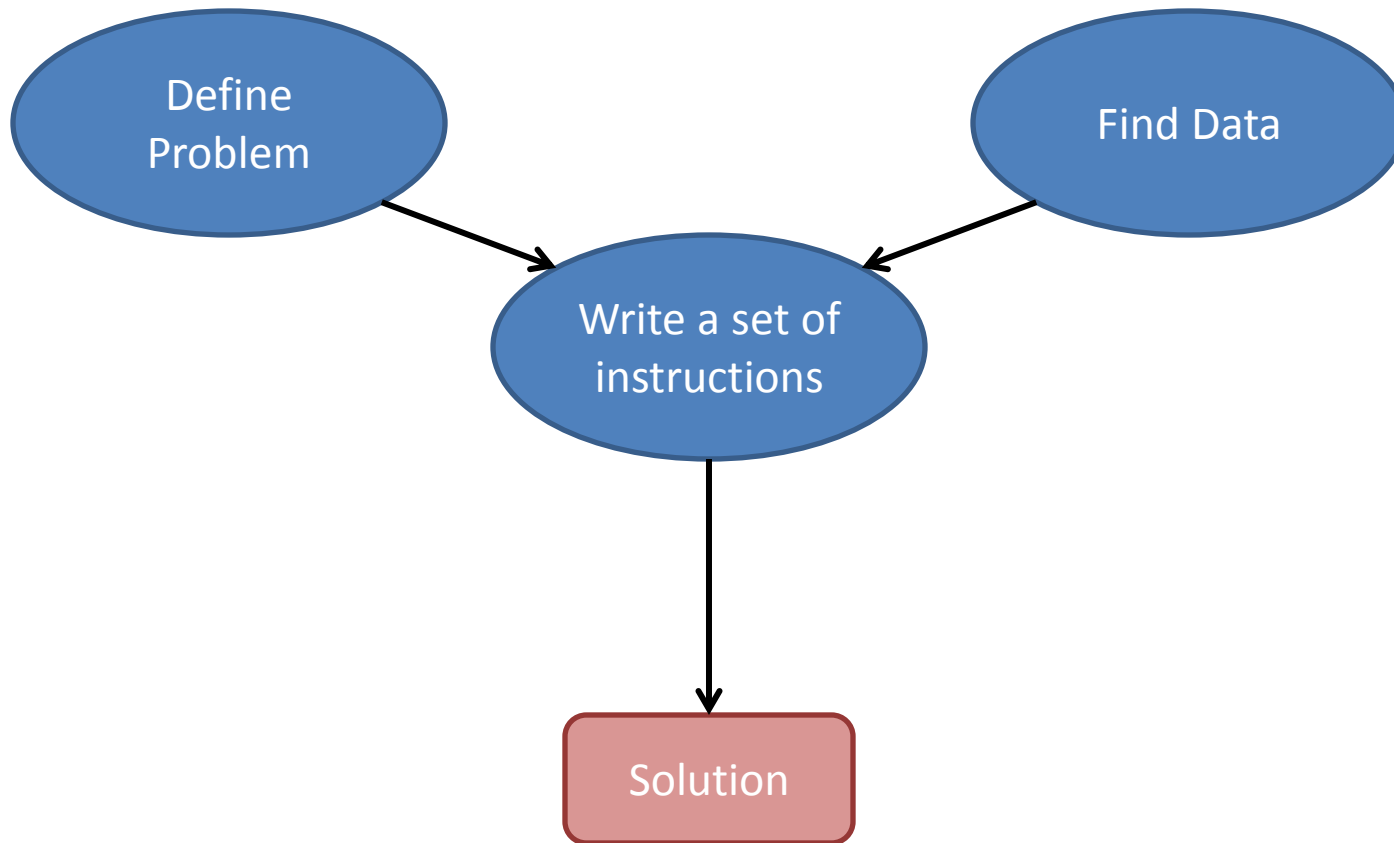
You will meet with Jadrian and discuss your project. Students will sign up for a time starting on May 7th and ending through the assigned Exam Period (Friday, May 10th from 9am to noon). This meeting has multiple goals:

1. Ensure that the course staff has access to all the submitted materials (i.e. you remembered to add all the files in the final handin).
2. Determine how well you understand your program.
3. Provide feedback and advice about your analysis and results.

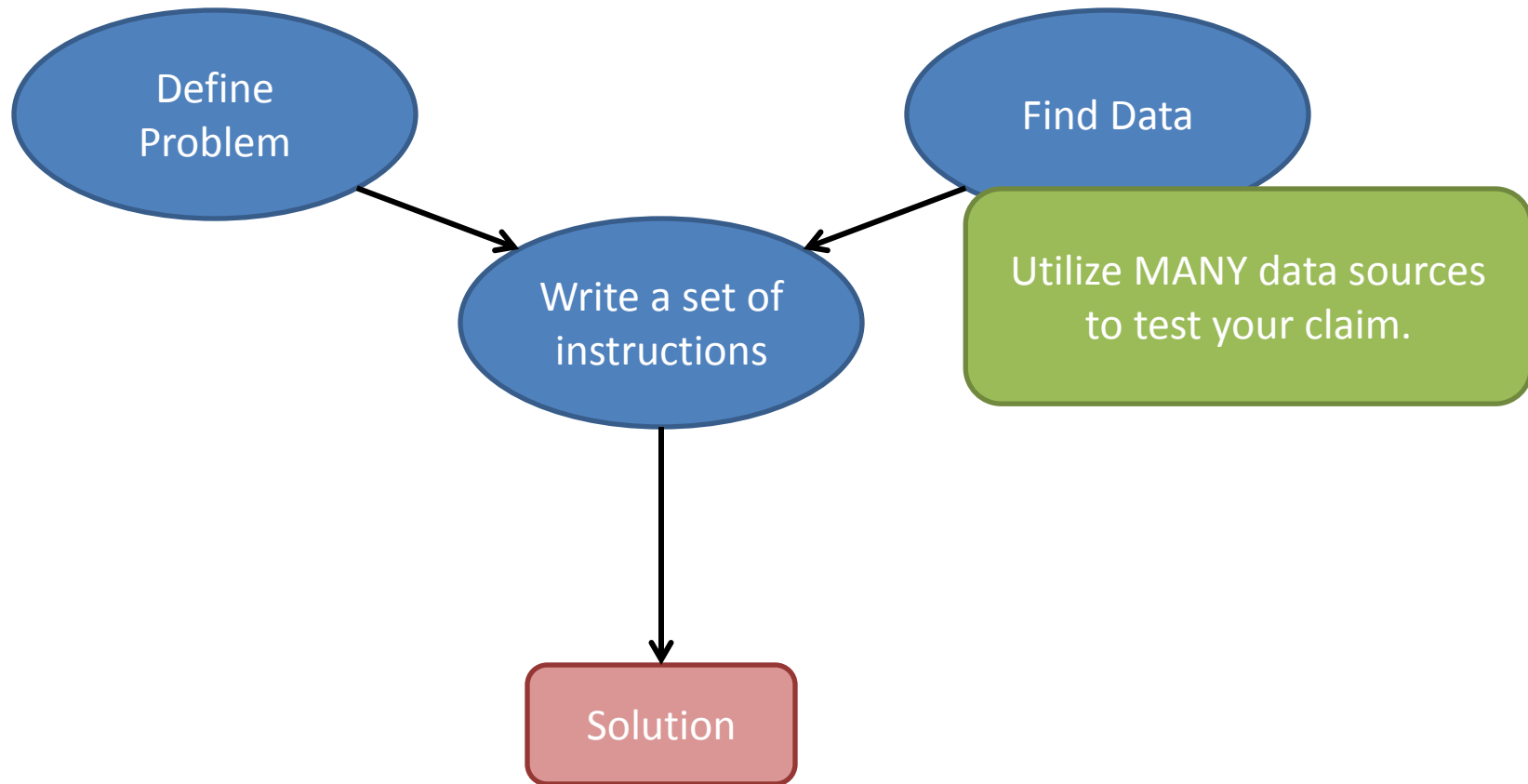
Today

- What do we wish we had known before?
- Example Final Projects
- Visitor!
- General Project Description
 - Scope
 - Timeline
- Let's Get Started

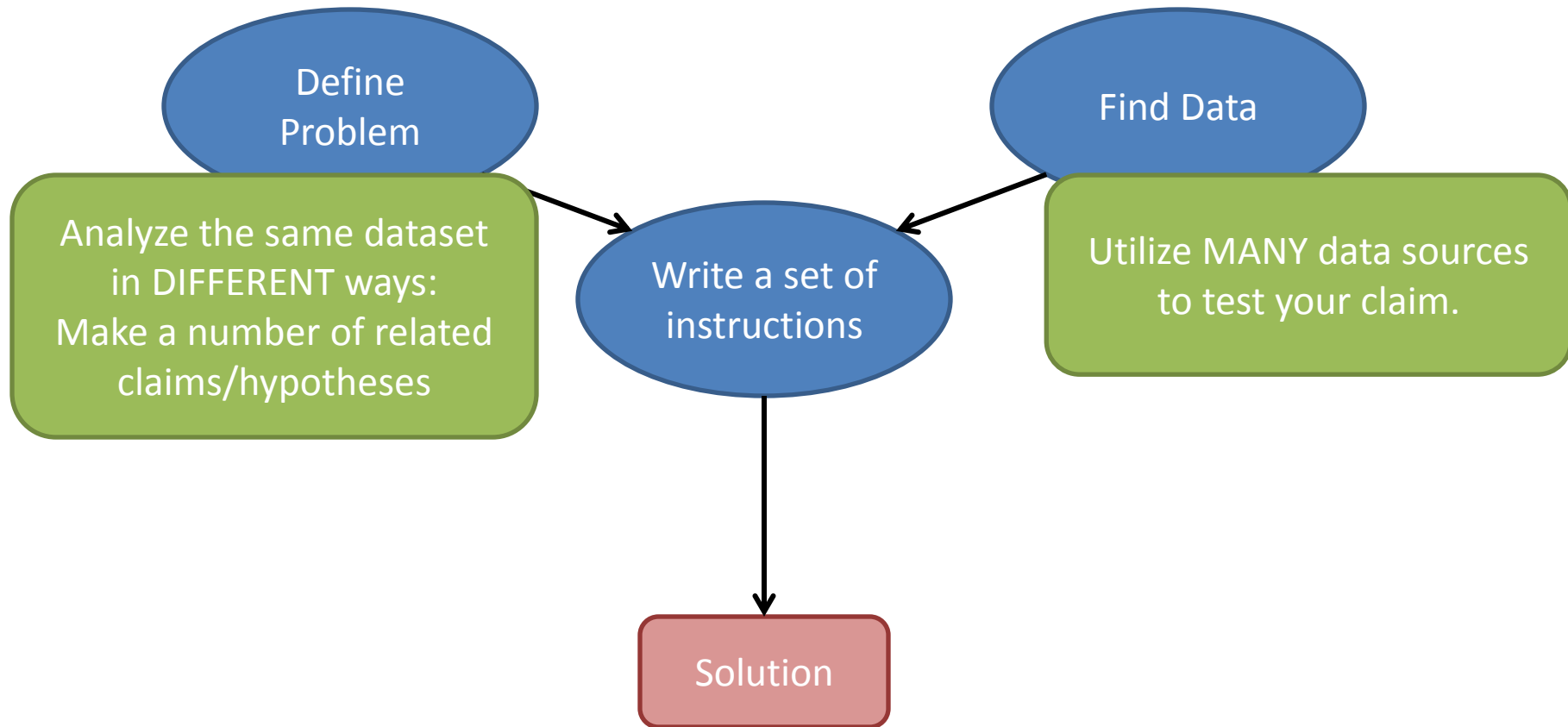
Scope – Think Big! (& Have Backup Plans)



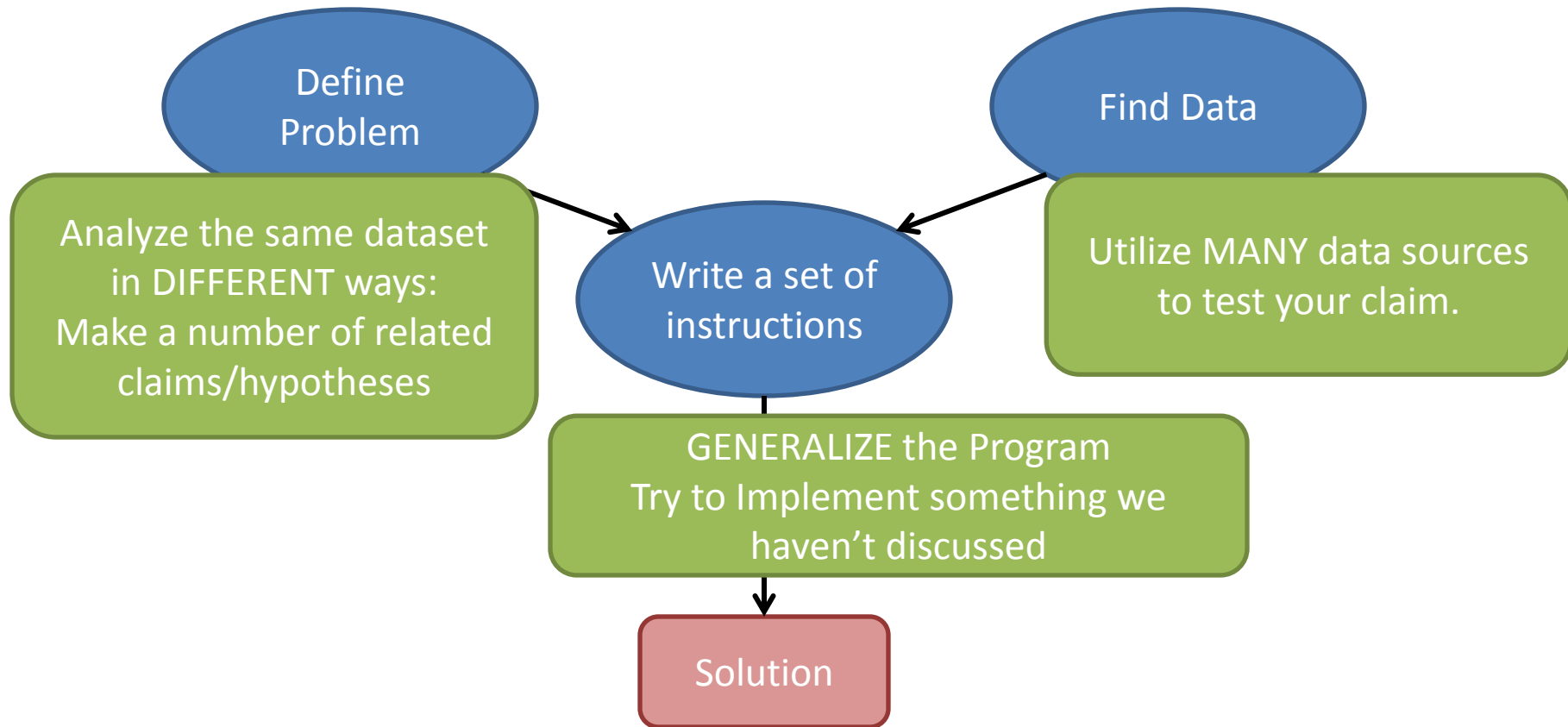
Scope – Think Big! (& Have Backup Plans)



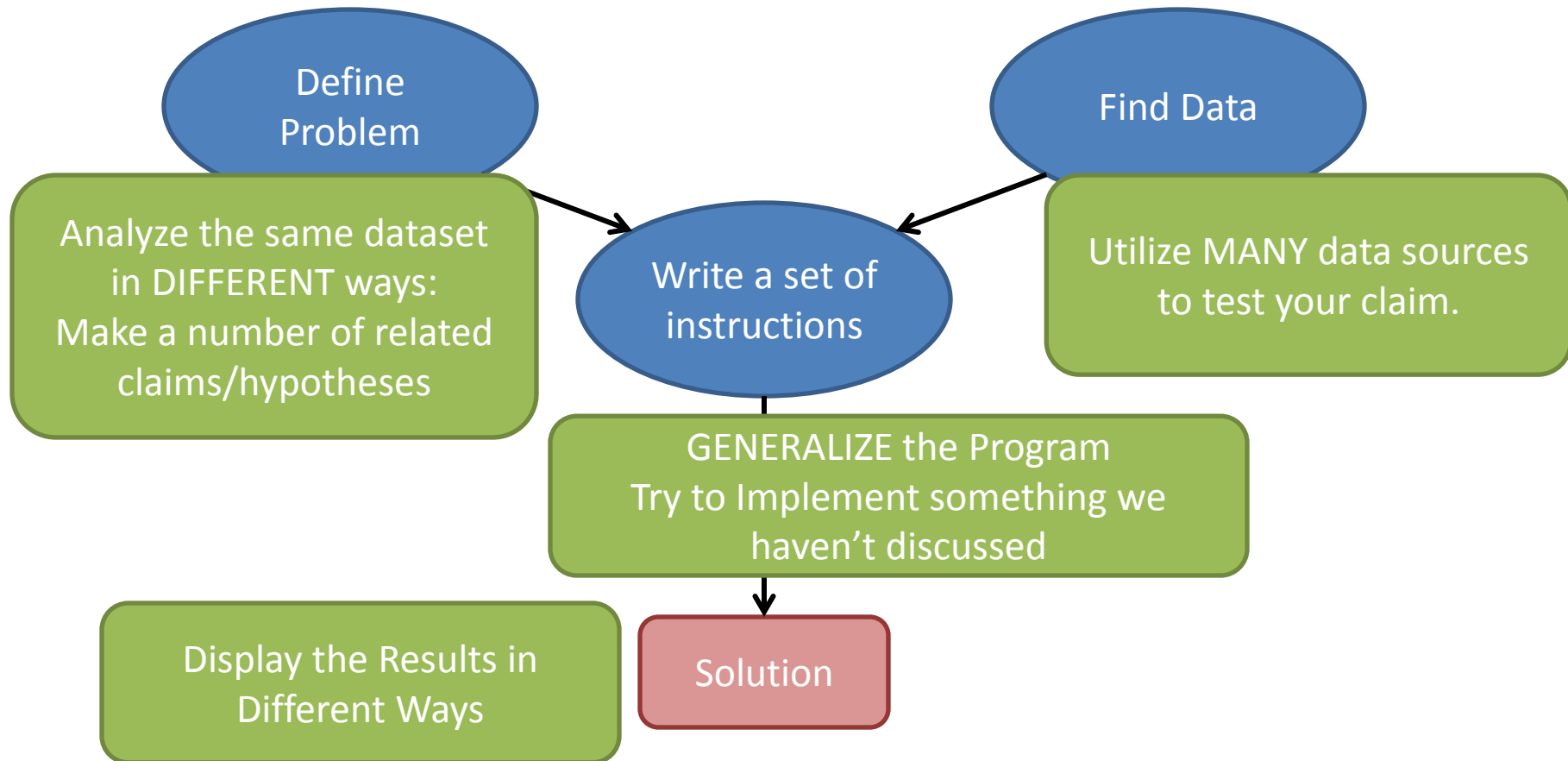
Scope – Think Big! (& Have Backup Plans)



Scope – Think Big! (& Have Backup Plans)



Scope – Think Big! (& Have Backup Plans)



Something New: Clustering

Something New: Clustering

Here's an Algorithm for Clustering Data into k groups:

1. Randomly select k data points.
2. Assign all other data points to the *closest* selected point.
3. Recompute each group's *mean* point.
4. Repeat Steps 2 & 3 until nothing changes any more.

Something New: Clustering

Here's an Algorithm for Clustering Data into k groups:

1. Randomly select k data points.
2. Assign all other data points to the *closest* selected point.
3. Recompute each group's *mean* point.
4. Repeat Steps 2 & 3 until nothing changes any more.

Called k -Means Clustering

Today

- What do we wish we had known before?
- Example Final Projects
- Visitor!
- General Project Description
 - Scope
 - Timeline
- **Let's Get Started**

Timeline

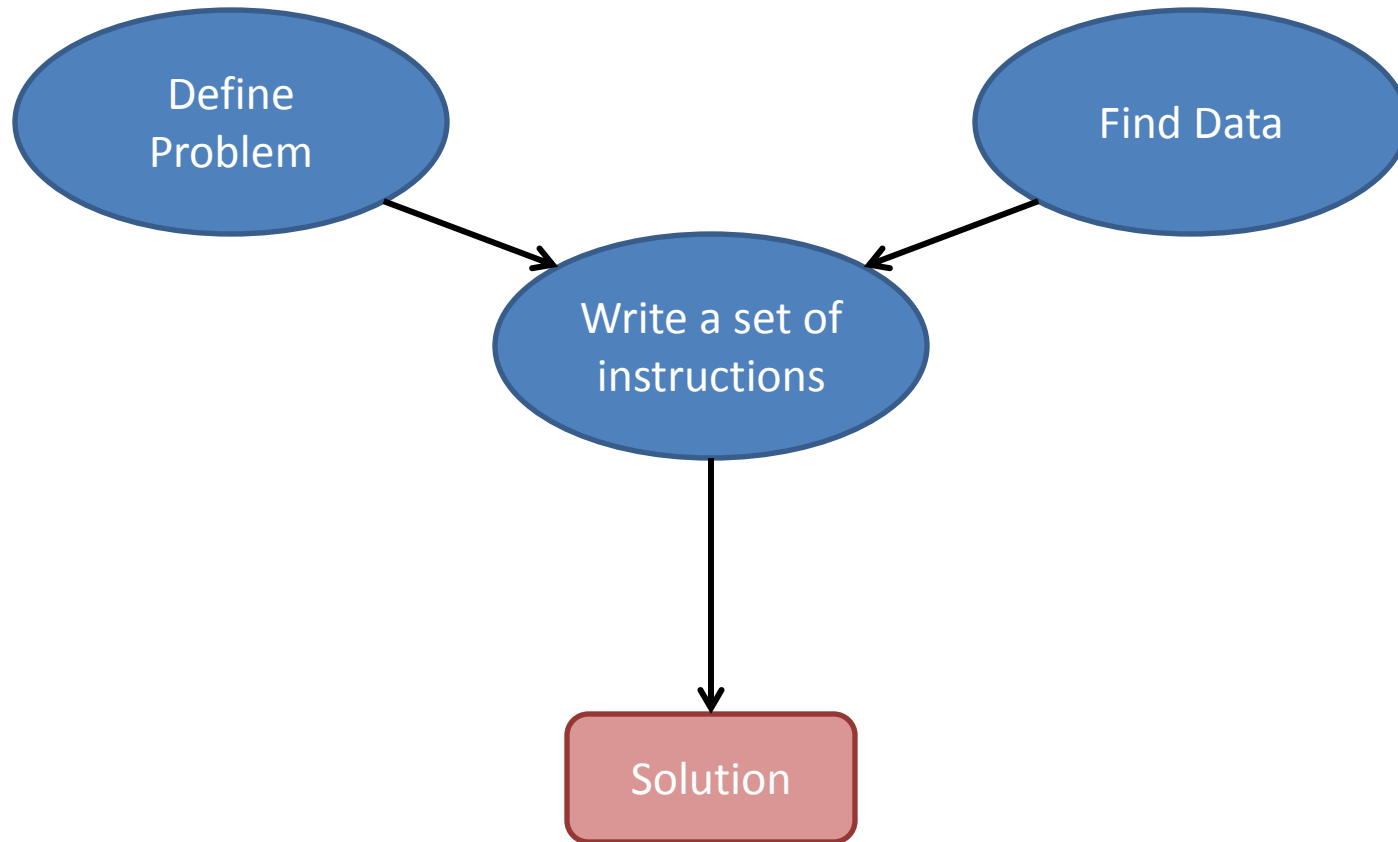
| Sun | Mon | Tues | Wed | Thurs | Fri | Sat |
|-----|-------------------------------|---|---------------------|--------------------|---------------------|-----|
| | 4/15 | 4/16 | 4/17 | 4/18 | 4/19 | |
| | | Final Project Out | | In-Class Work Time | | |
| | 4/22 | 4/23 | 4/24 | 4/25 | 4/26 | |
| | Meet With Staff At Least Once | | | | | |
| | | In-Class Work Time Proposal Due | | In-Class Work Time | | |
| | 4/29 | 4/30 | 5/1 | 5/2 | 5/3 | |
| | Meet With Staff At Least Once | | | | | |
| | | In-Class Work Time | | In-Class Work Time | | |
| | 5/6 | 5/7 | 5/8 | 5/9 | 5/10 | |
| | | Final Project Due | Interactive Grading | | | |
| | | | | | Exam Period 9-12 | |

Today

- What do we wish we had known before?
- Example Final Projects
- Visitor!
- General Project Description
 - Scope
 - Timeline
- **Let's Get Started**

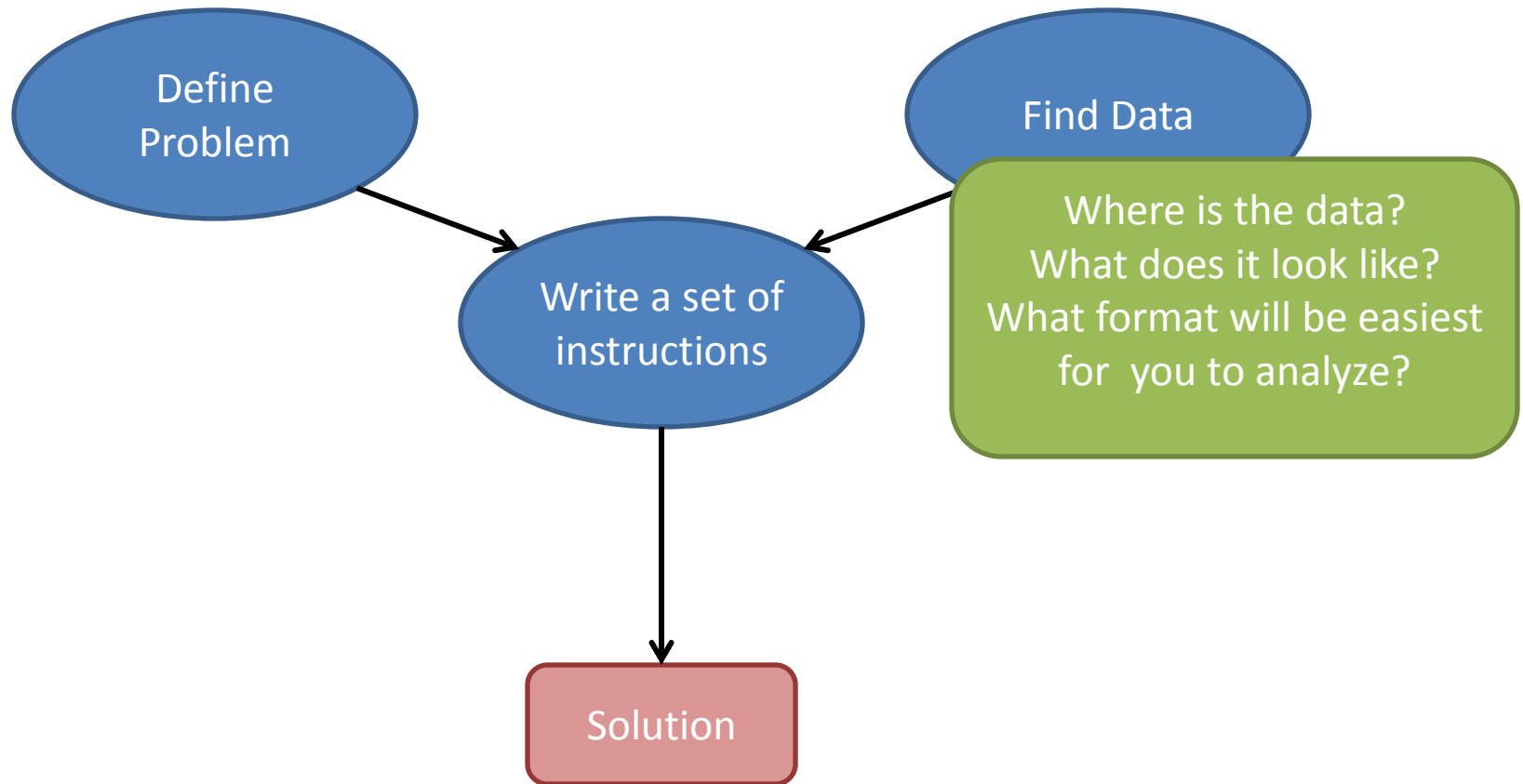
Before You Begin...

Say you have a vague idea of what you want to do....



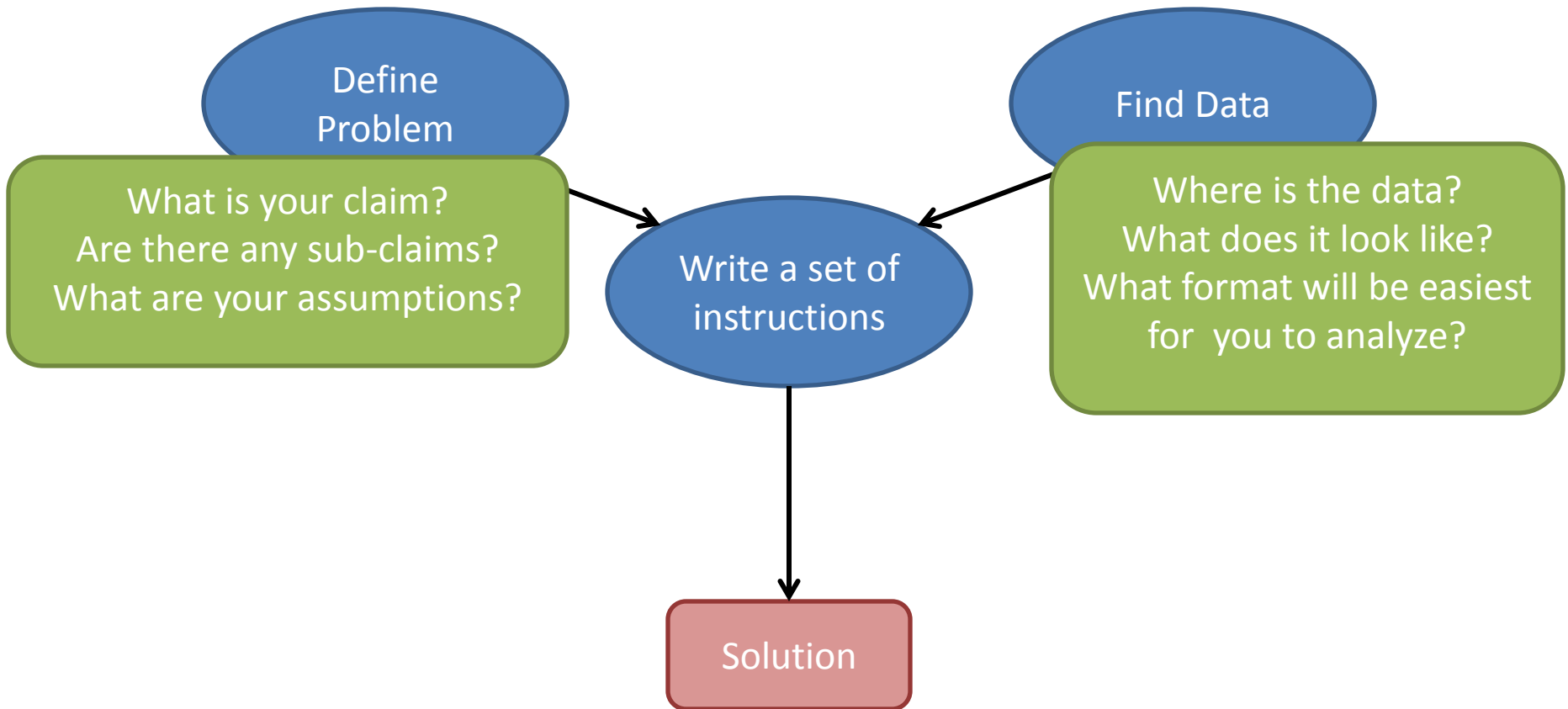
Before You Begin...

Say you have a vague idea of what you want to do....



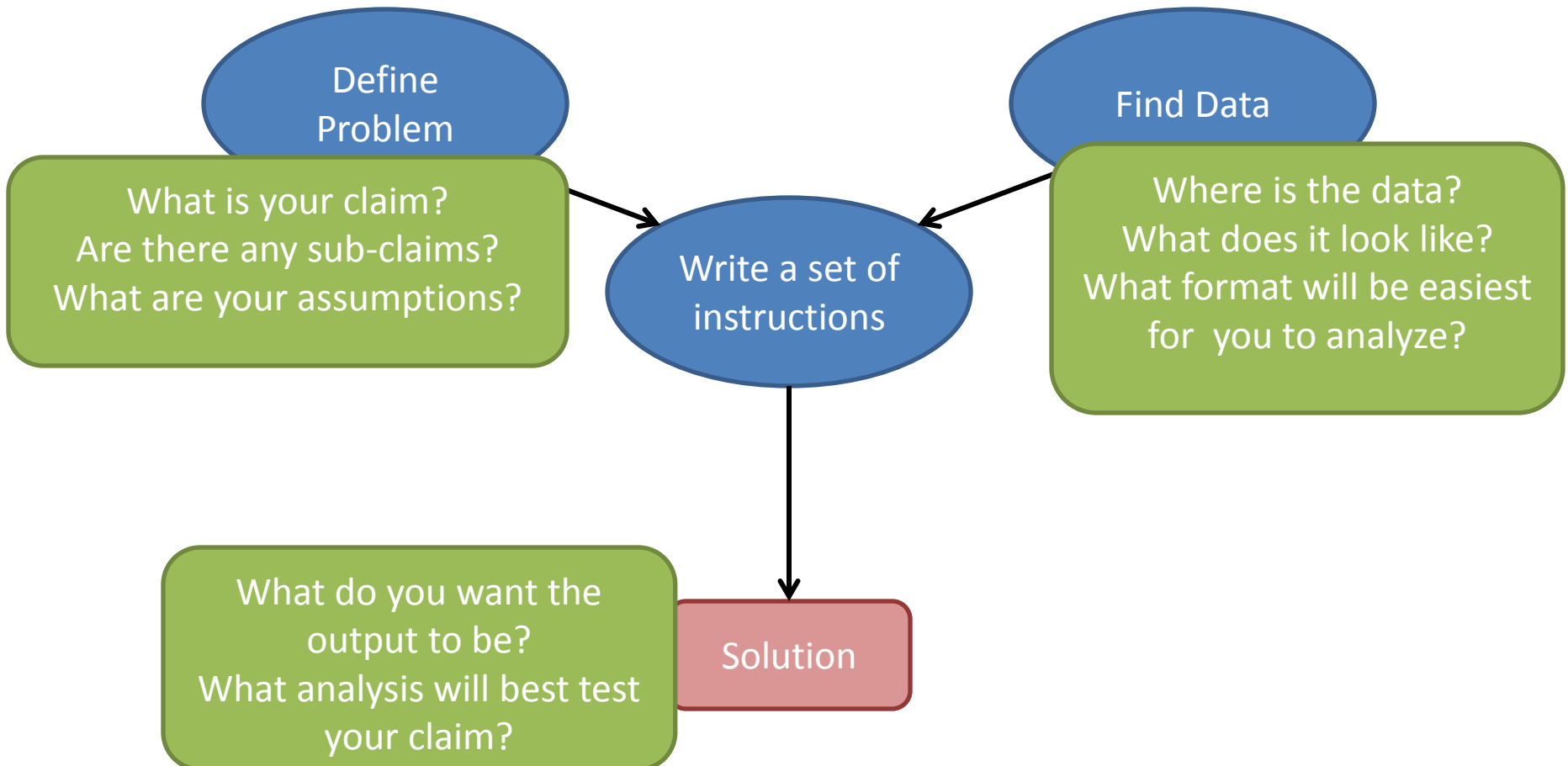
Before You Begin...

Say you have a vague idea of what you want to do....



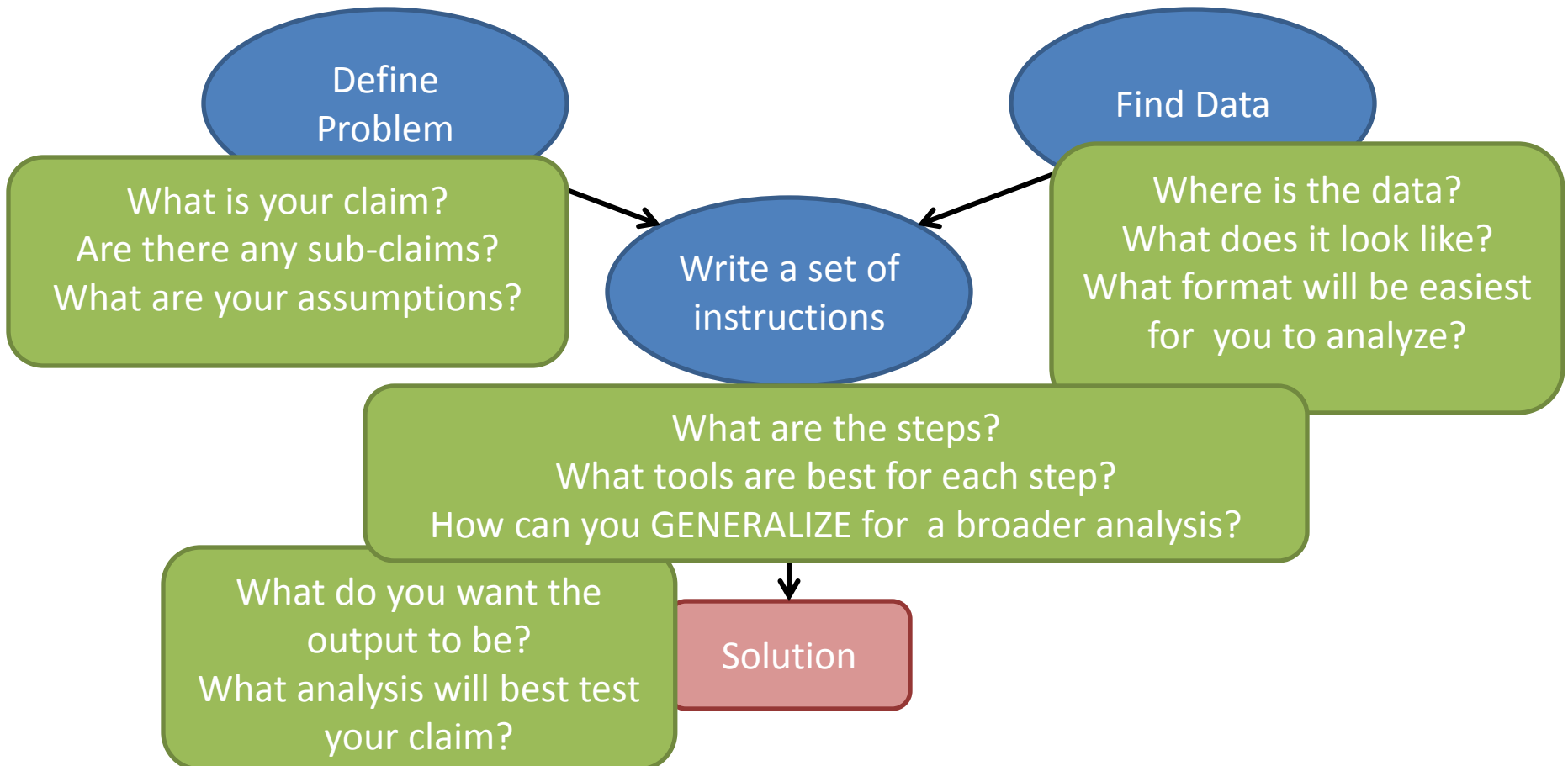
Before You Begin...

Say you have a vague idea of what you want to do....



Before You Begin...

Say you have a vague idea of what you want to do....



Let's Get Started!

- Pitch an Idea to a Partner
- Tell the Class your Idea

Before You Begin...

