

Betty Harper, Shannon Lee, and Carly Sunseri

The Pennsylvania State University

Build it right the first time: Modeling survey data for BI tools

Welcome!
We are...



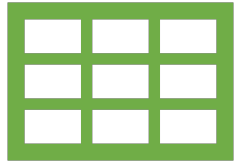
Betty Harper, Assistant Director of Assessment

Shannon Lee, Research Associate

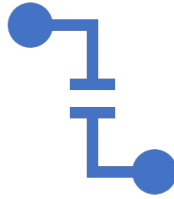
Carly Sunseri, Data and Reporting Analyst

The Office of Planning, Assessment, and
Institutional Research (OPAIR), The Pennsylvania
State University

By the end of this session, you should be able to:



Describe flat files
and their
limitations



Sketch out a
basic relational
data model



List the key steps
to build a data
model



Find additional
resources for
self-guided
learning

...so that you can get started on your survey data visualization journey

Session Roadmap

Flat files and their limitations

Facts and dimensions

Building your base files in Excel

Building the data model

Conclusion



Building Blocks

The initial steps in your journey

Flat files

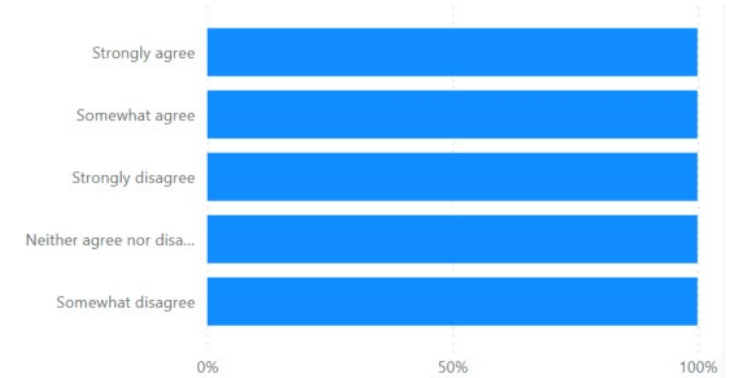
The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	I	J	K	L	M	
1	UniqueID	Q1	Q2	Q8	Q9	Q10	Q11	Q12	Q13
2	1	Strongly disagree	Strongly disagree	Often	Poor	Heck yeah!	Man	18 - 24	White
3	2	Strongly disagree	Neither agree nor disagree	Never	Very good	Not sure	Woman	55 - 64	White
4	3	Strongly disagree	Strongly disagree	Never	Poor	Heck yeah!	Man	18 - 24	White
5	4	Strongly disagree	Strongly disagree	Never	Good	Are you kidding? No way!	Man		Unknown
6	5	Strongly disagree		Never	Very good	Not sure	Woman	35 - 44	White
7	6	Strongly disagree	Somewhat disagree	Never	Fair	Not sure		25 - 34	Unknown
8	7	Strongly disagree	Strongly disagree	Never	Very good	Heck yeah!	Woman	45 - 54	White
9	8	Strongly disagree	Strongly agree	Never	Very good	Heck yeah!	Man	45 - 54	White
10	9	Strongly disagree	Strongly disagree	Never	Fair	Are you kidding? No way!	Woman	55 - 64	White
11	10	Strongly disagree	Somewhat agree	Never	Good	Are you kidding? No way!	Woman	18 - 24	White

	A	B	C	D	E	F
1	Unique	Q1	Q2	Q3	Q4	Q5
2	1	Strongly disagree	Strongly disagree	Somewhat disagree	Strongly disagree	Strongly disagree
3	2	Strongly disagree	Neither agree nor disagree	Strongly disagree	Strongly disagree	Somewhat disagree
4	3	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
5	4	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree	Somewhat disagree
6	5	Strongly disagree		Strongly disagree	Strongly disagree	Strongly disagree
7	6	Strongly disagree	Somewhat disagree	Strongly disagree	Somewhat agree	Strongly disagree
8	7	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
9	8	Strongly disagree	Strongly agree	Strongly agree	Strongly agree	Strongly agree
10	9	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
11	10	Strongly disagree	Somewhat agree	Strongly agree	Strongly agree	Somewhat agree
12	11	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
13	12	Strongly disagree		Strongly agree	Strongly agree	Strongly agree
14	13	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
15	14	Strongly disagree	Somewhat agree	Somewhat agree	Somewhat agree	Strongly disagree
16	15	Strongly disagree		Strongly disagree	Strongly disagree	Strongly disagree
17	16	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
18	17	Strongly disagree		Somewhat disagree	Strongly disagree	Neither agree nor disagree
19	18	Strongly disagree		Strongly disagree	Strongly disagree	Strongly agree

RawDataSet

- Q1
- Q10
- Q11
- Q12
- Q13
- Q2
- Q3
- Q4
- Q5
- Q6
- Q7



Why a Data Model?

Facts vs. dimensions

Facts = things that can be measured such as multiple-choice questions

To what extent do you agree with the following statements?	
Q1	I feel welcome at Faber College.
Q2	I feel valued at Faber College.
Q3	I feel safe at Faber College.
Q4	I think fraternities are an important part of Faber College student life.
Q5	I think togas are appropriate classroom attire.
Q6	I do not appreciate the pop culture references in this survey.
Please answer the following questions.	
Q7	How often do you skip class to attend fraternity parties?
Q8	How often do you take advantage of Dean Wormer's office hours?
Q9	How would you rate the overall quality of your education Faber College?
Q10	Would you recommend Faber College to others?

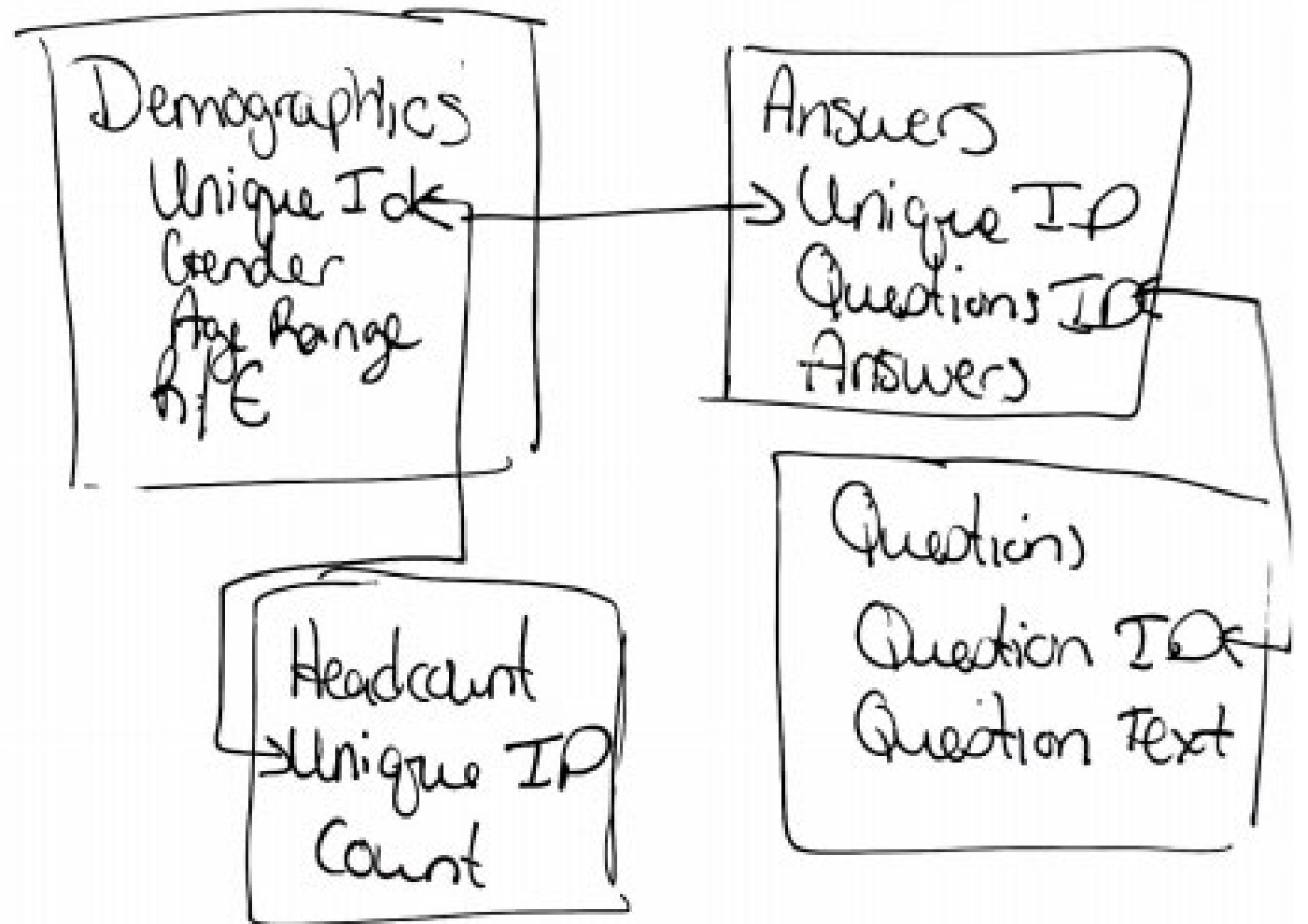
Dimensions = companion data such as demographics

Q11	Gender identity
Q12	Age range
Q13	IPEDS Race

Where do I start?

Fields

- Unique ID
- Q1 – Q10
- Gender
- Age Range
- Race/Ethnicity



A close-up, shallow depth-of-field photograph of a computer keyboard. The central focus is a dark grey key with a white 'X' symbol. To its right, a key with a white vertical bar is partially visible. The background shows other keys, including a blue arrow key, all blurred. The lighting is soft, creating a professional and clean aesthetic.

Creating your base files in Excel

- Answers
- Question Key
- Headcount
- Demographics
- Value Order

*Your survey data may result in
more files than this

Survey answers (facts)

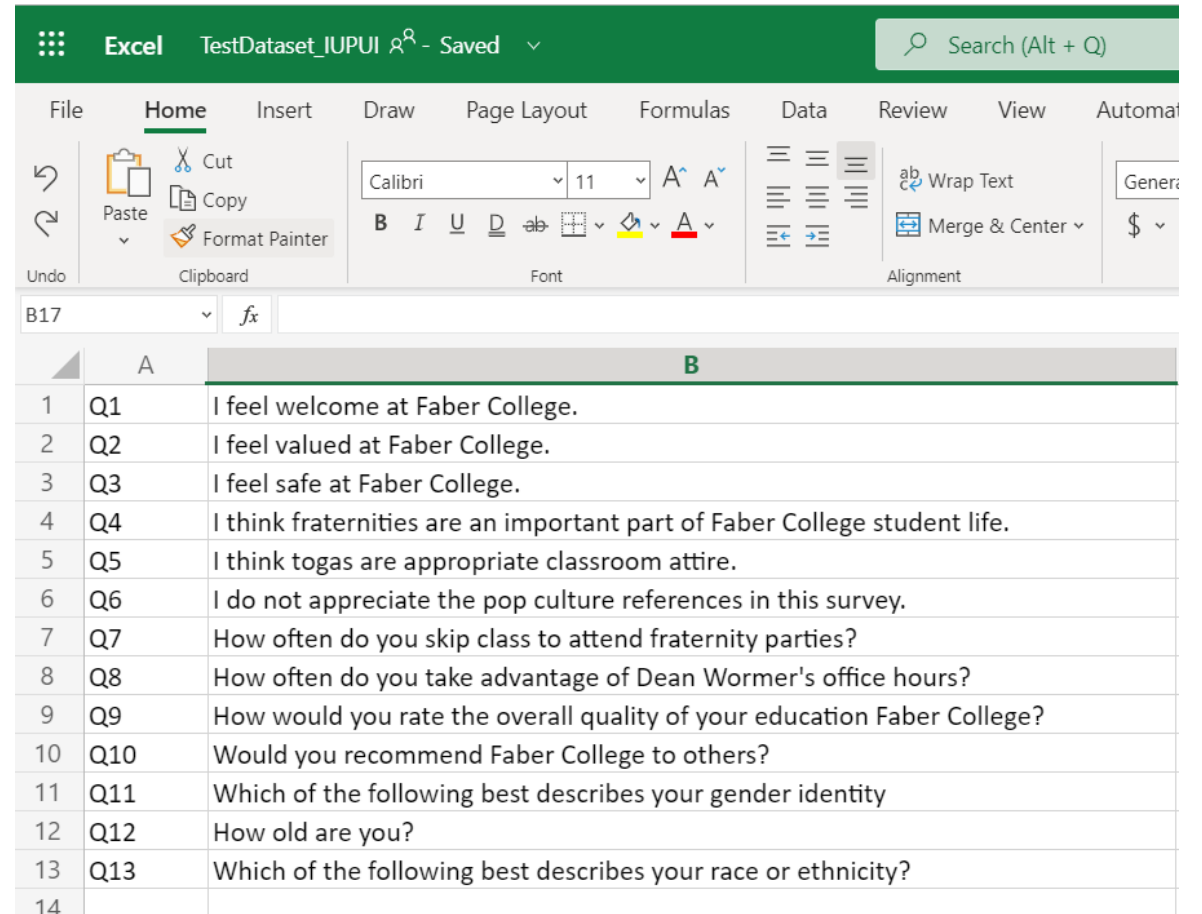
Excel TestDataset_IUPUI A - Saved

File Home Insert Draw Page Layout Formulas Data Review View Automate Help Open in Desktop App

Undo Paste Copy Format Painter Clipboard Font Alignment Number Tables

	A	B	C	D	E
1	UniqueID	Q1	Q2	Q3	Q4
2	1	Strongly disagree	Strongly disagree	Somewhat disagree	Strongly disagree
3	2	Strongly disagree	Neither agree nor disagree	Strongly disagree	Strongly disagree
4	3	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
5	4	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
6	5	Strongly disagree		Strongly disagree	Strongly disagree
7	6	Strongly disagree	Somewhat disagree	Strongly disagree	Somewhat agree
8	7	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
9	8	Strongly disagree	Strongly agree	Strongly agree	Strongly agree
10	9	Strongly disagree	Strongly disagree	Strongly disagree	Strongly disagree
11	10	Strongly disagree	Somewhat agree	Strongly agree	Strongly agree

The question key



Excel TestDataset_IUPUI - Saved

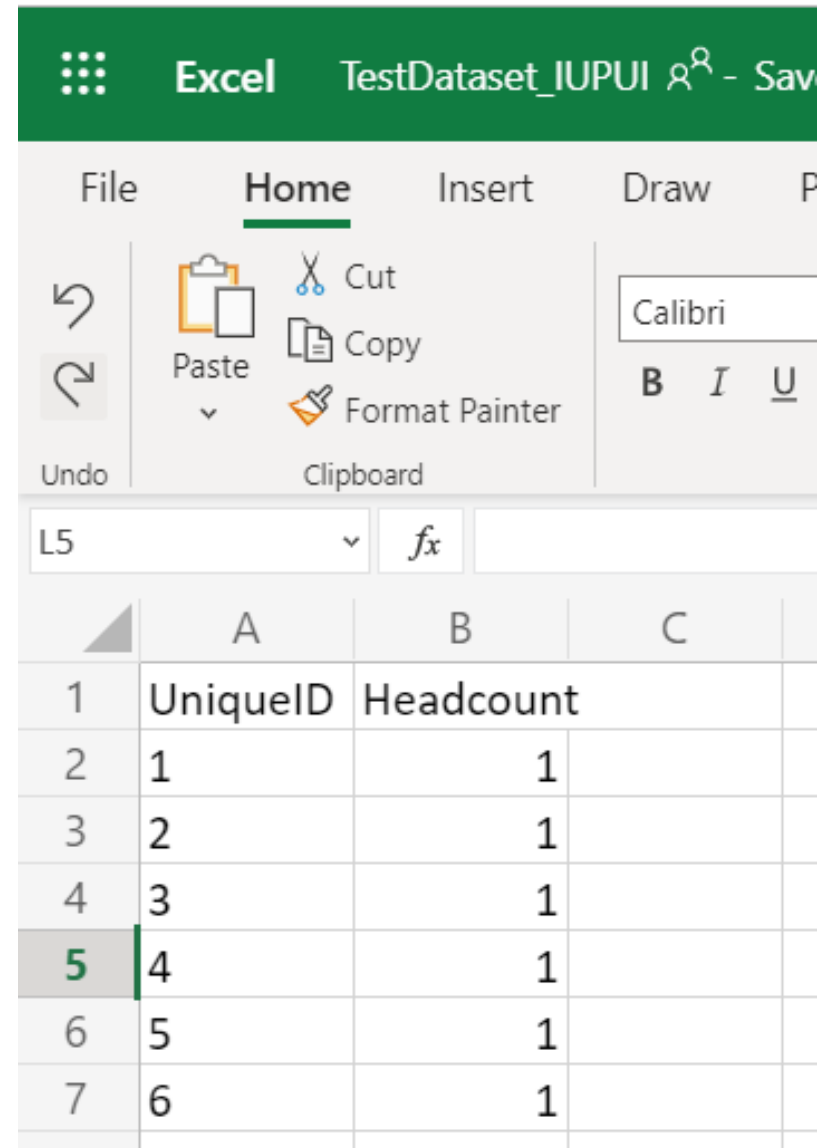
File Home Insert Draw Page Layout Formulas Data Review View Automate

Undo Paste Copy Format Painter Clipboard Font Alignment

B17

	A	B
1	Q1	I feel welcome at Faber College.
2	Q2	I feel valued at Faber College.
3	Q3	I feel safe at Faber College.
4	Q4	I think fraternities are an important part of Faber College student life.
5	Q5	I think togas are appropriate classroom attire.
6	Q6	I do not appreciate the pop culture references in this survey.
7	Q7	How often do you skip class to attend fraternity parties?
8	Q8	How often do you take advantage of Dean Wormer's office hours?
9	Q9	How would you rate the overall quality of your education Faber College?
10	Q10	Would you recommend Faber College to others?
11	Q11	Which of the following best describes your gender identity
12	Q12	How old are you?
13	Q13	Which of the following best describes your race or ethnicity?
14		

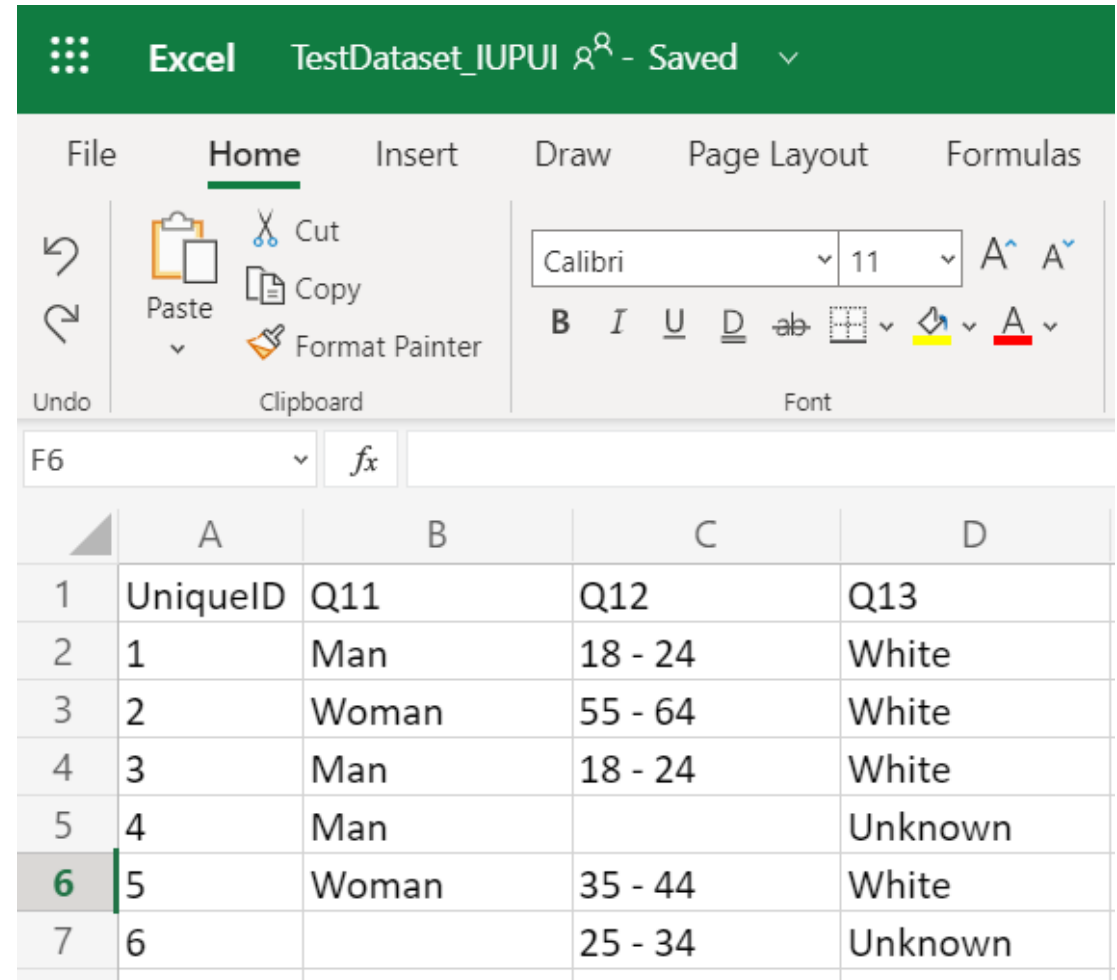
Headcount



The screenshot shows the Microsoft Excel interface. The title bar indicates the file name is 'TestDataset_IUPUI 88 - Save'. The 'Home' tab is active in the ribbon, showing options for Undo, Paste, Cut, Copy, and Format Painter. The font is set to Calibri, and the font size is 11. The spreadsheet has two columns: 'UniqueID' and 'Headcount'. The data is as follows:

	A	B	C
1	UniqueID	Headcount	
2	1	1	
3	2	1	
4	3	1	
5	4	1	
6	5	1	
7	6	1	

Demographics (dimensions)



Excel TestDataset_IUPUI 8 - Saved

File Home Insert Draw Page Layout Formulas

Undo Paste Copy Format Painter Clipboard

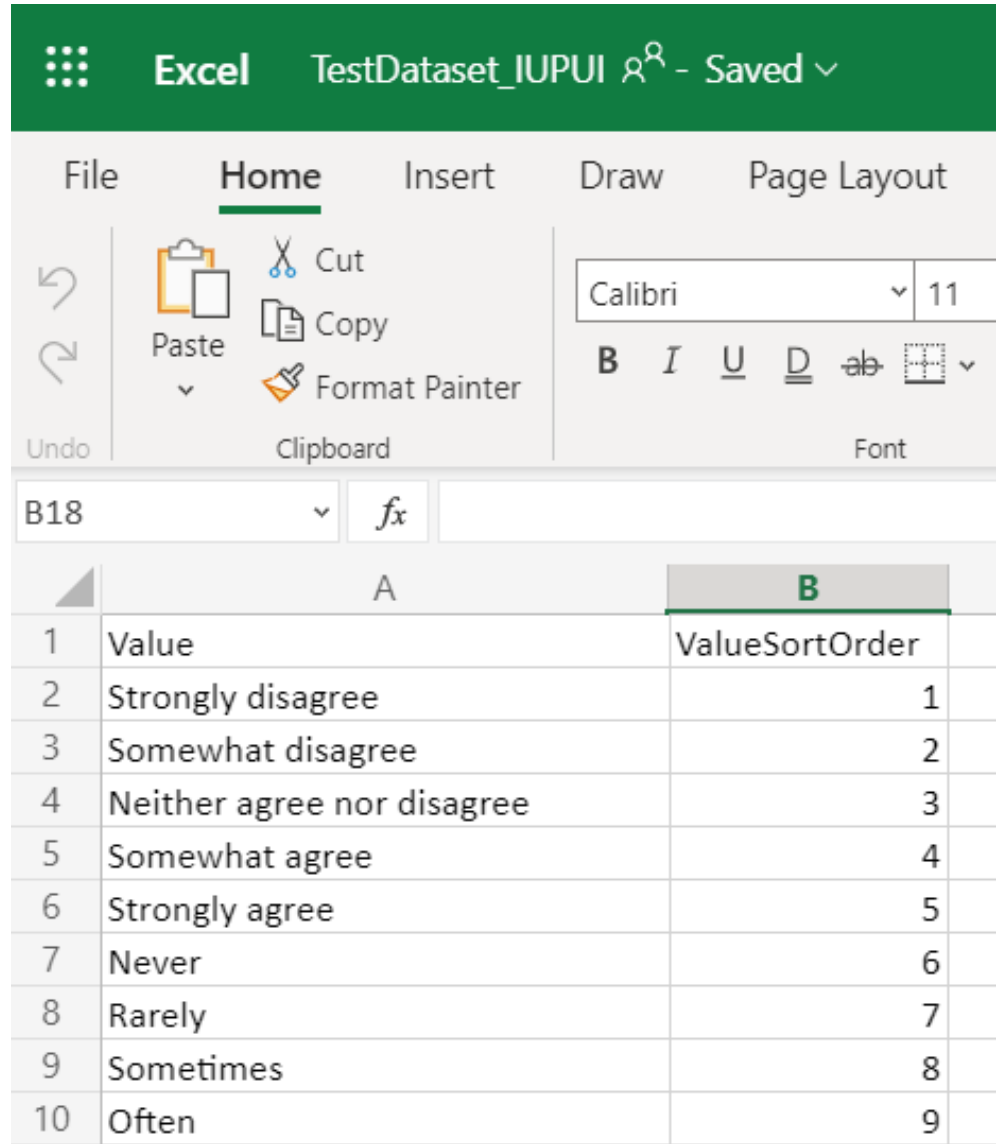
Font Calibri 11 A^ A^ B I U D ab [Grid] [Color] A^

F6 fx

	A	B	C	D
1	UniqueID	Q11	Q12	Q13
2	1	Man	18 - 24	White
3	2	Woman	55 - 64	White
4	3	Man	18 - 24	White
5	4	Man		Unknown
6	5	Woman	35 - 44	White
7	6		25 - 34	Unknown

Value Order

(used to make sure responses are presented in your desired order later)



Excel TestDataset_IUPUI - Saved

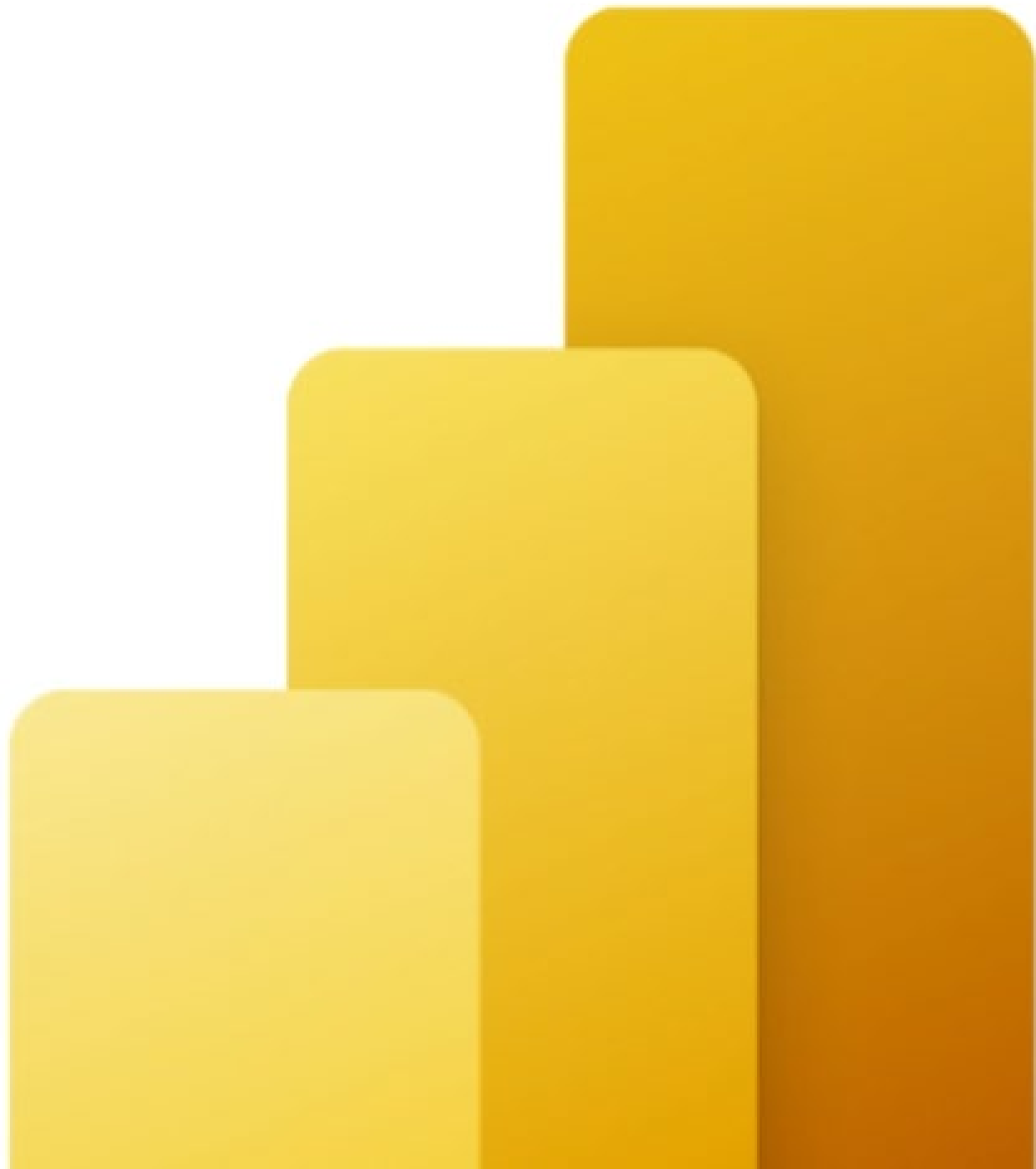
File Home Insert Draw Page Layout

Undo Paste Cut Copy Format Painter

Clipboard Font

B18 fx

	A	B
1	Value	ValueSortOrder
2	Strongly disagree	1
3	Somewhat disagree	2
4	Neither agree nor disagree	3
5	Somewhat agree	4
6	Strongly agree	5
7	Never	6
8	Rarely	7
9	Sometimes	8
10	Often	9

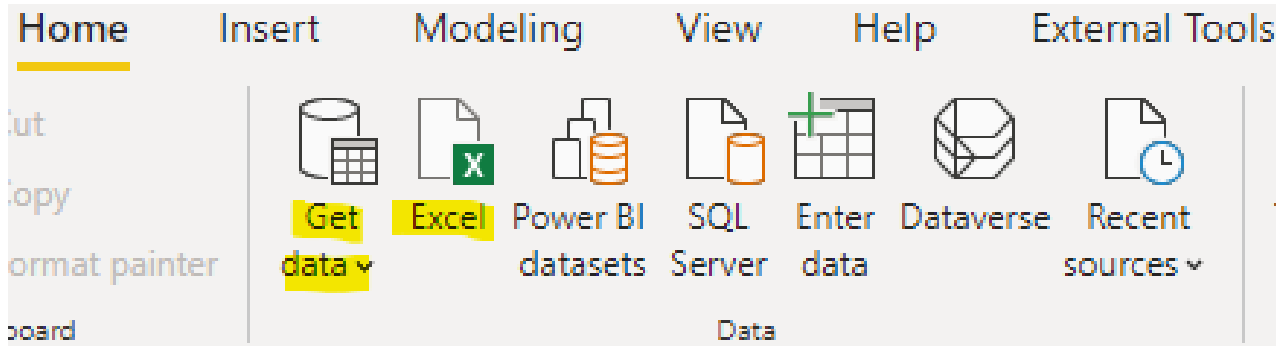


Importing your data into your visualization tool

We will use Power BI to demonstrate

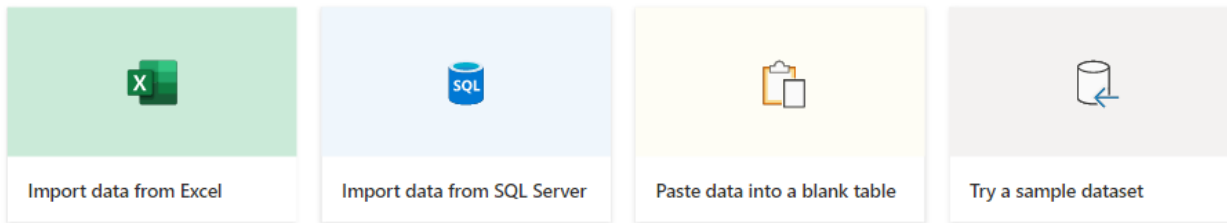
Importing Your Data

Power BI makes it easy!



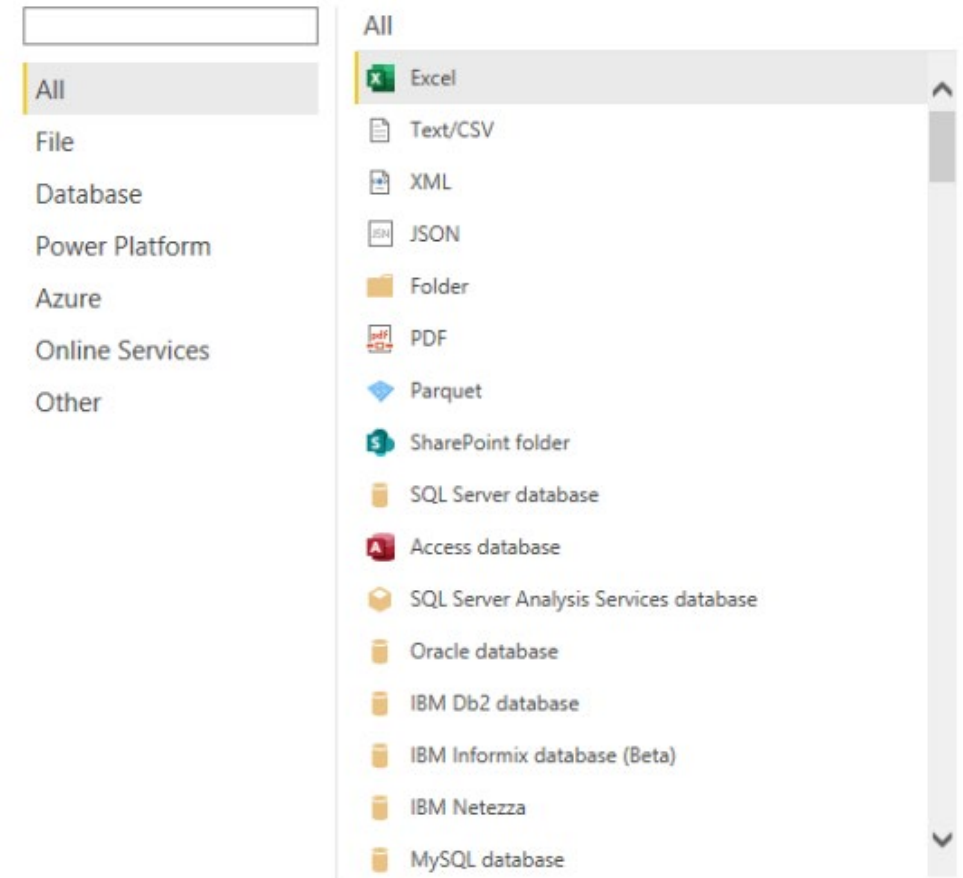
Add data to your report

Once loaded, your data will appear in the **Fields** pane.



[Get data from another source →](#)

Get Data



Certified Connectors

Template Apps

Connect

Cancel

Loading Your Data

Navigator

Display Options

TestDataset_IUPUI.xlsx [6]

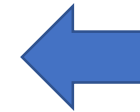
- Answers
- Demographics
- Headcount
- QuestionKey
- RawDataSet
- ValueOrder



These are the individual spreadsheets from the workbook

Answers

UniqueID	Q1	Q2	Q3
1	Strongly disagree	Strongly disagree	Somewhat disagree
2	Strongly disagree	Neither agree nor disagree	Strongly disagree
3	Strongly disagree	Strongly disagree	Strongly disagree
4	Strongly disagree	Strongly disagree	Strongly disagree
5	Strongly disagree		null Strongly disagree
6	Strongly disagree	Somewhat disagree	Strongly disagree
7	Strongly disagree	Strongly disagree	Strongly disagree
8	Strongly disagree	Strongly agree	Strongly agree
9	Strongly disagree	Strongly disagree	Strongly disagree
10	Strongly disagree	Somewhat agree	Strongly agree
11	Strongly disagree	Strongly disagree	Strongly disagree
12	Strongly disagree		null
13	Strongly disagree	Strongly disagree	Strongly disagree
14	Strongly disagree	Somewhat agree	Somewhat agree
15	Strongly disagree		null Strongly disagree
16	Strongly disagree	Strongly disagree	Strongly disagree
17	Strongly disagree		null Somewhat disagree
18	Strongly disagree		null Strongly disagree
19	Strongly disagree	Strongly disagree	
20	Strongly disagree	Strongly disagree	Strongly disagree
21	Strongly disagree	Strongly disagree	Strongly disagree
22	Strongly disagree	Strongly disagree	Strongly disagree



You can preview your data before loading it

Option to transform data prior to loading (can transform after as well)

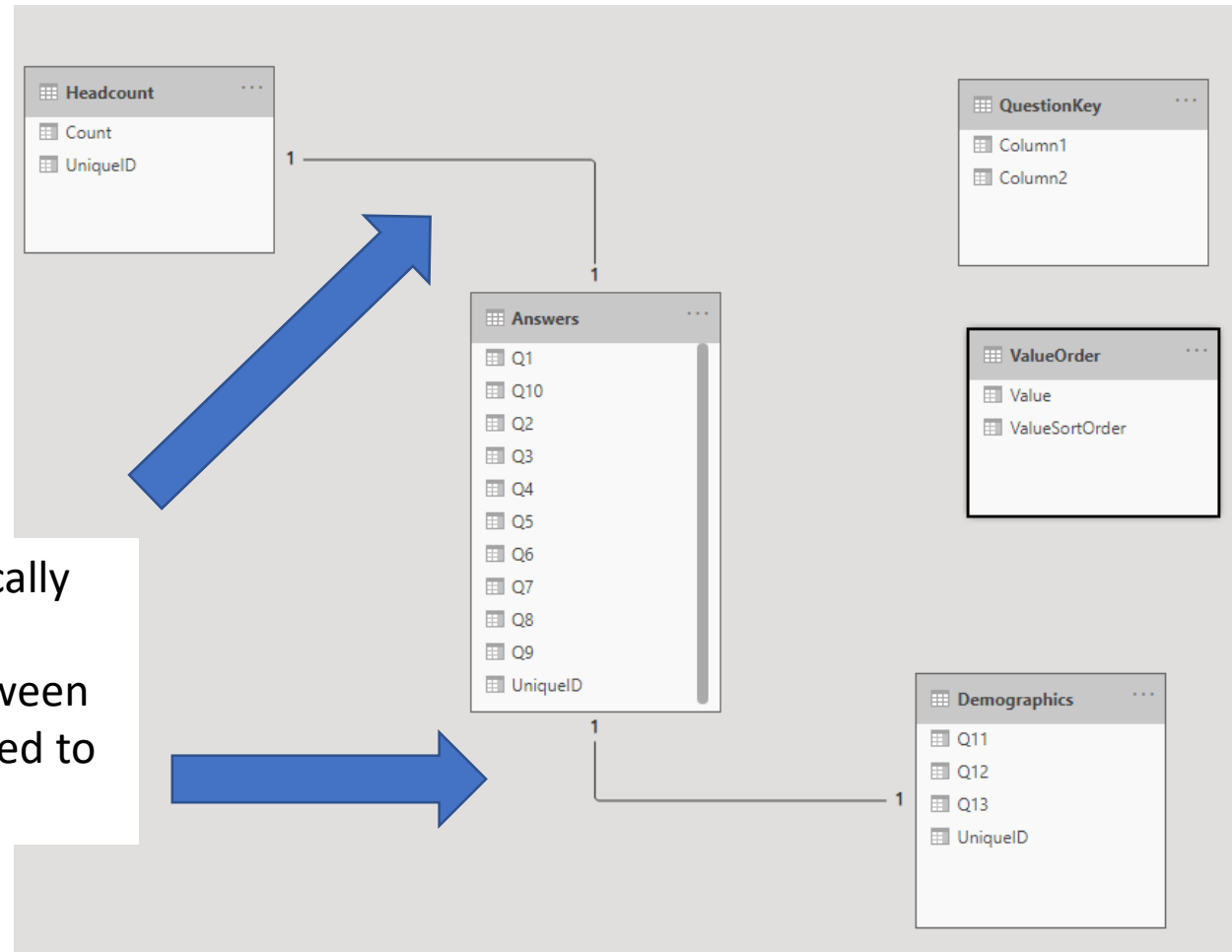
Load Transform Data Cancel



Creating your data model

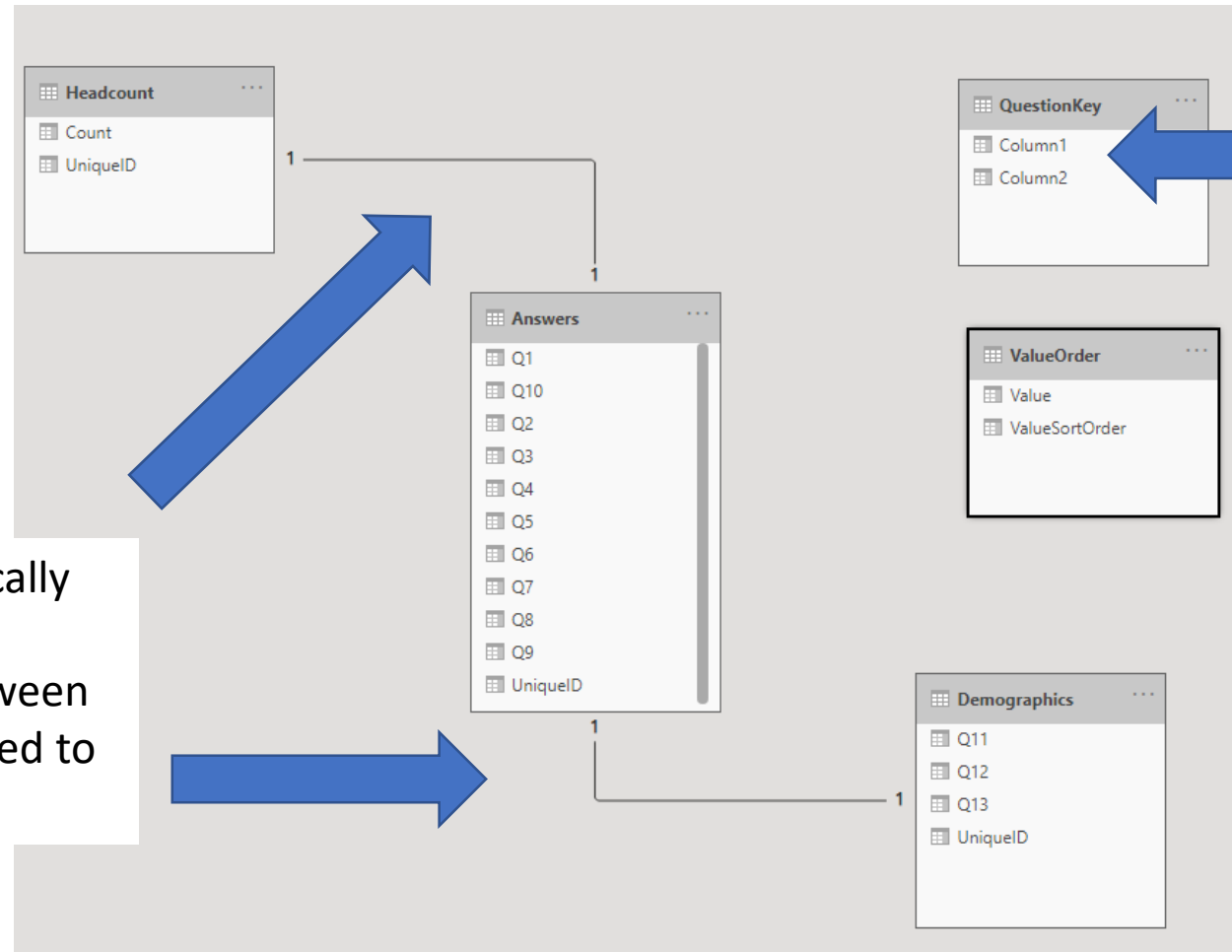
Again, we will use Power BI to demonstrate

Review how the data were imported



Power BI is automatically assuming there are relationships between the tables, but we need to specify what they are

Review how the data were imported

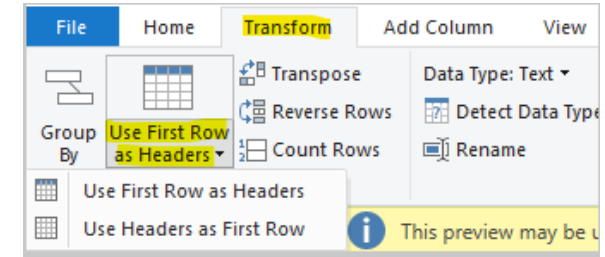


Some column names may not import cleanly, but you can easily fix this

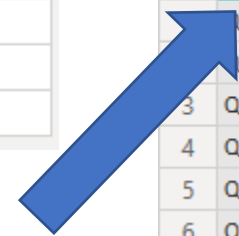
Power BI is automatically assuming there are relationships between the tables, but we need to specify what they are

Promote Column Headers

	Column1	Column2
1	QuestionID	QuestionText
2	Q1	I feel welcome at Faber College.
3	Q2	I feel valued at Faber College.
4	Q3	I feel safe at Faber College.
5	Q4	I think fraternities are an important part of Faber College student life.
6	Q5	I think togas are appropriate classroom attire.
7	Q6	I do not appreciate the pop culture references in this survey.
8	Q7	How often do you skip class to attend fraternity parties?
9	Q8	How often do you take advantage of Dean Wormer's office hours?
10	Q9	How would you rate the overall quality of your education Faber Colleg...
11	Q10	Would you recommend Faber College to others?
12	Q11	Which of the following best describes your gender identity
13	Q12	How old are you?
14	Q13	Which of the following best describes your race or ethnicity?



	QuestionID	QuestionText
1		I feel welcome at Faber College.
2		I feel valued at Faber College.
3	Q3	I feel safe at Faber College.
4	Q4	I think fraternities are an important part of Faber College student life.
5	Q5	I think togas are appropriate classroom attire.
6	Q6	I do not appreciate the pop culture references in this survey.
7	Q7	How often do you skip class to attend fraternity parties?
8	Q8	How often do you take advantage of Dean Wormer's office hours?
9	Q9	How would you rate the overall quality of your education Faber Colleg...
10	Q10	Would you recommend Faber College to others?
11	Q11	Which of the following best describes your gender identity
12	Q12	How old are you?
13	Q13	Which of the following best describes your race or ethnicity?

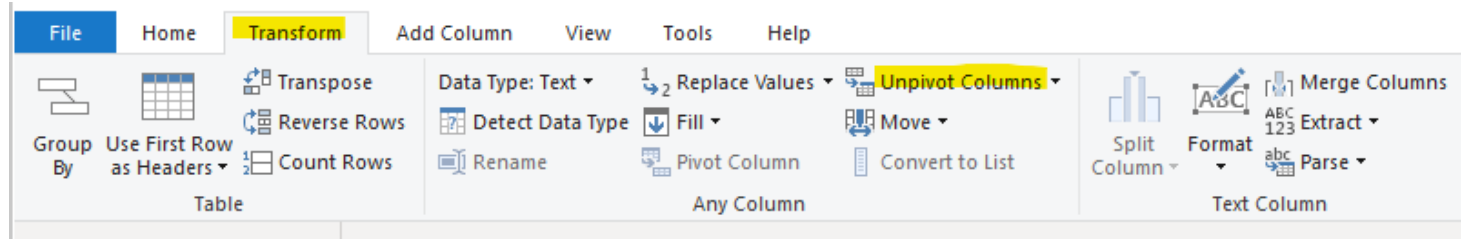
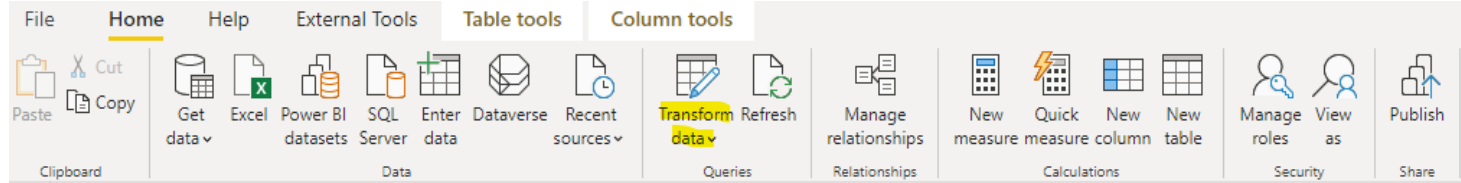


Make sure to fix any typos :)

To do this in Power BI, double click the column title.

Unpivot

Open PowerQuery



You'll want to rename this to QuestionID



Wide Format

UniquelD	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
53	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Never	Never	Very good	Are you kidding? No way!



	UniquelD	Attribute	Value
467	53	Q1	Strongly agree
468	53	Q2	Strongly agree
469	53	Q3	Strongly agree
470	53	Q4	Strongly agree
471	53	Q5	Strongly agree
472	53	Q6	Strongly agree
473	53	Q7	Never
474	53	Q8	Never
475	53	Q9	Very good
476	53	Q10	Are you kidding? No way!

Long Format

Relationships

One-to-one (1:1) - the record only connects to one other record

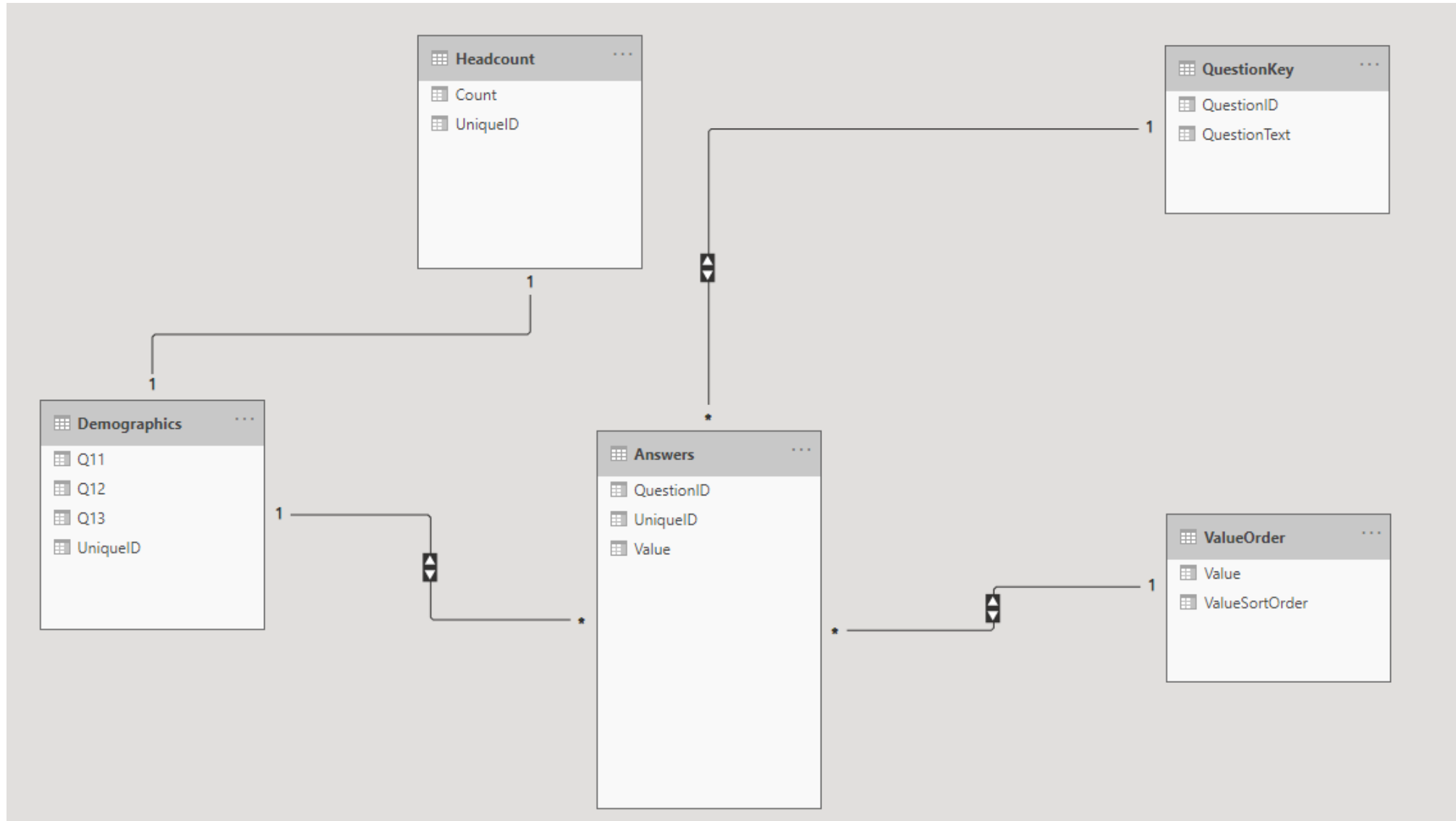
One-to-many (1:*) - one record can connect to many records

ex. A sports team: one team, multiple players.

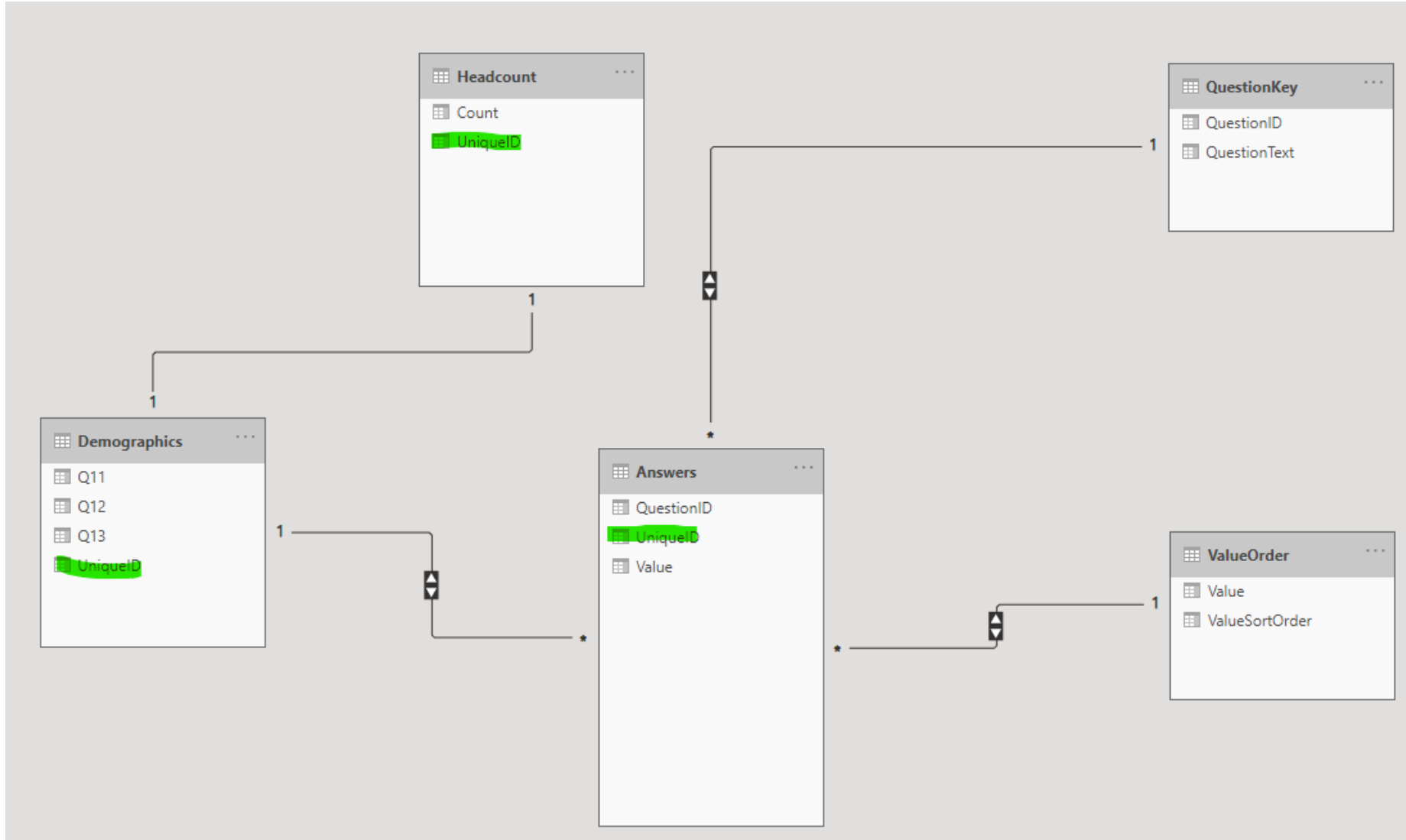
Many-to-many (*:*) - each record can connect to multiple records

ex. Students to classes for a semester. **Each class** has **multiple students** enrolled and **each student** can be enrolled in **multiple classes**

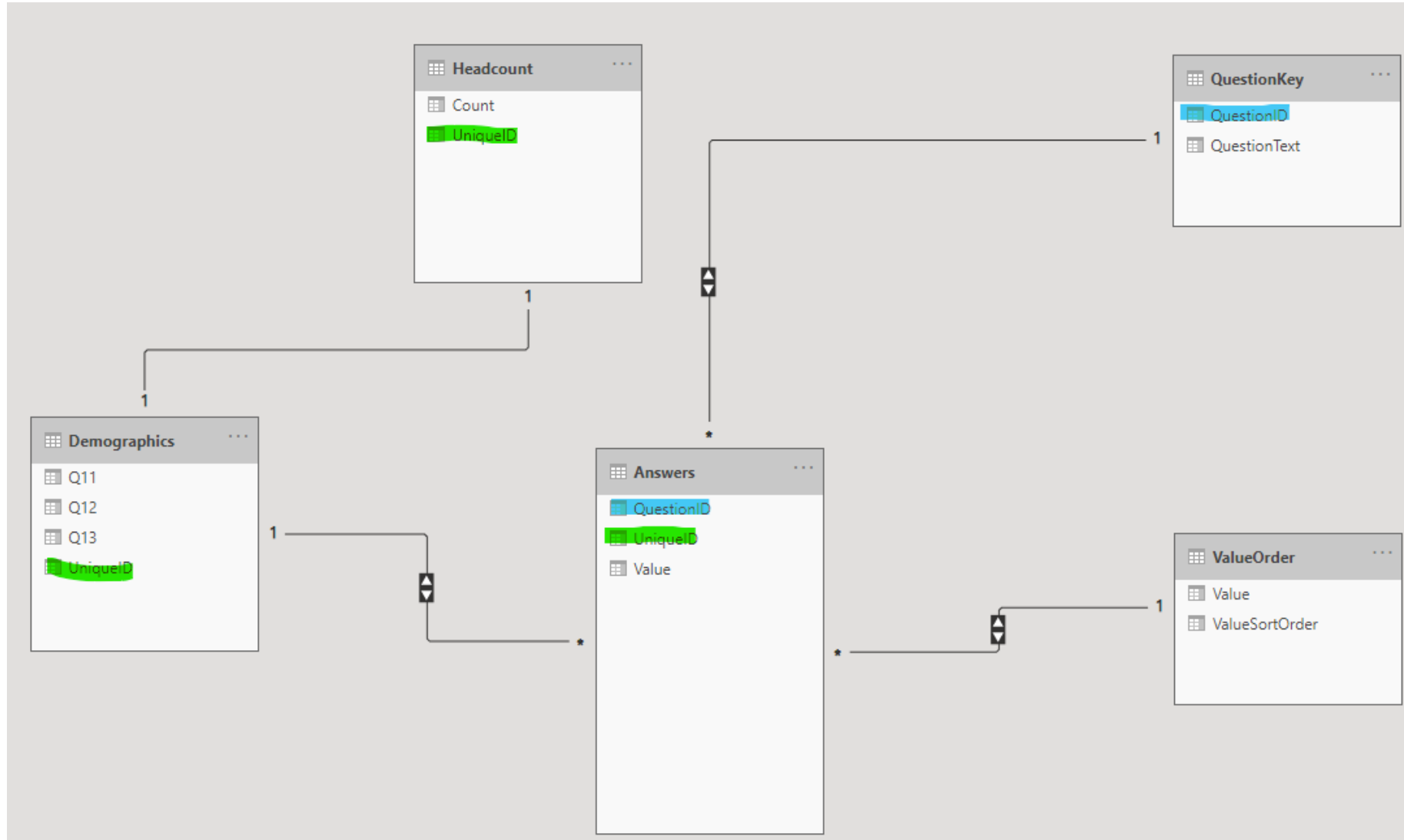
Establish Relationships



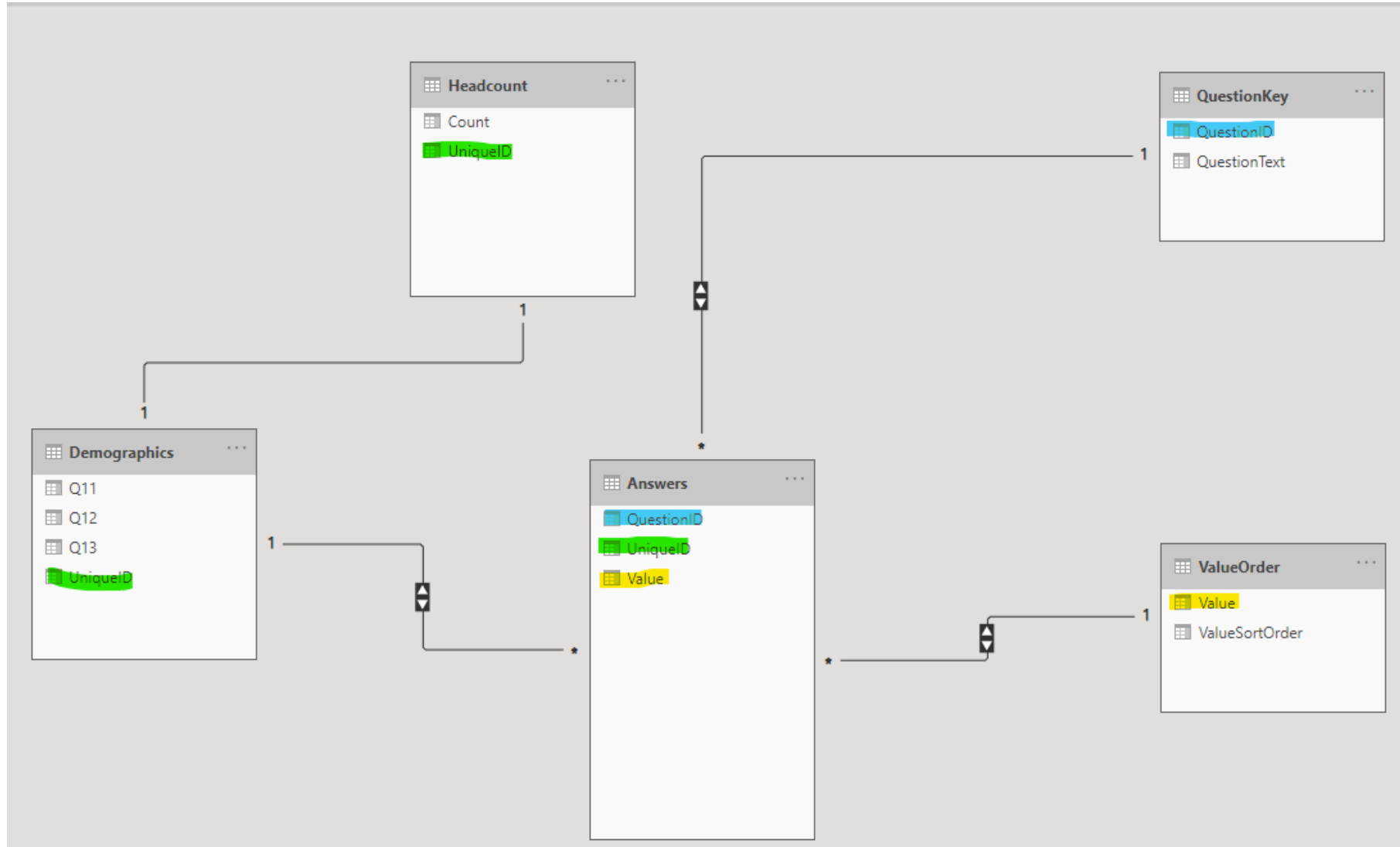
Establish Relationships



Establish Relationships



Establish Relationships



Manage & Edit Relationships

Manage relationships

Active	From: Table (Column)	To: Table (Column)
<input checked="" type="checkbox"/>	Answers (QuestionID)	QuestionKey (QuestionID)
<input checked="" type="checkbox"/>	Answers (UniqueID)	Demographics (UniqueID)
<input checked="" type="checkbox"/>	Answers (Value)	ValueOrder (Value)
<input checked="" type="checkbox"/>	Demographics (UniqueID)	Headcount (UniqueID)

New... Autodetect... Edit... Delete

Edit relationship

Select tables and columns that are related.

Answers

UniqueID	QuestionID	Value
53	Q1	Strongly agree
54	Q1	Strongly agree
55	Q1	Strongly agree

QuestionKey

QuestionID	QuestionText
Q1	I feel welcome at Faber College.
Q2	I feel valued at Faber College.
Q3	I feel safe at Faber College.

Cardinality

Many to one (*:1)

Cross filter direction

Both

Make this relationship active

Apply security filter in both directions

Assume referential integrity

OK

Cancel

FILTERS

Campus Group and Unit
All

Role at the University
All

Citizenship
All

Race/Ethnicity
All

Gender
All

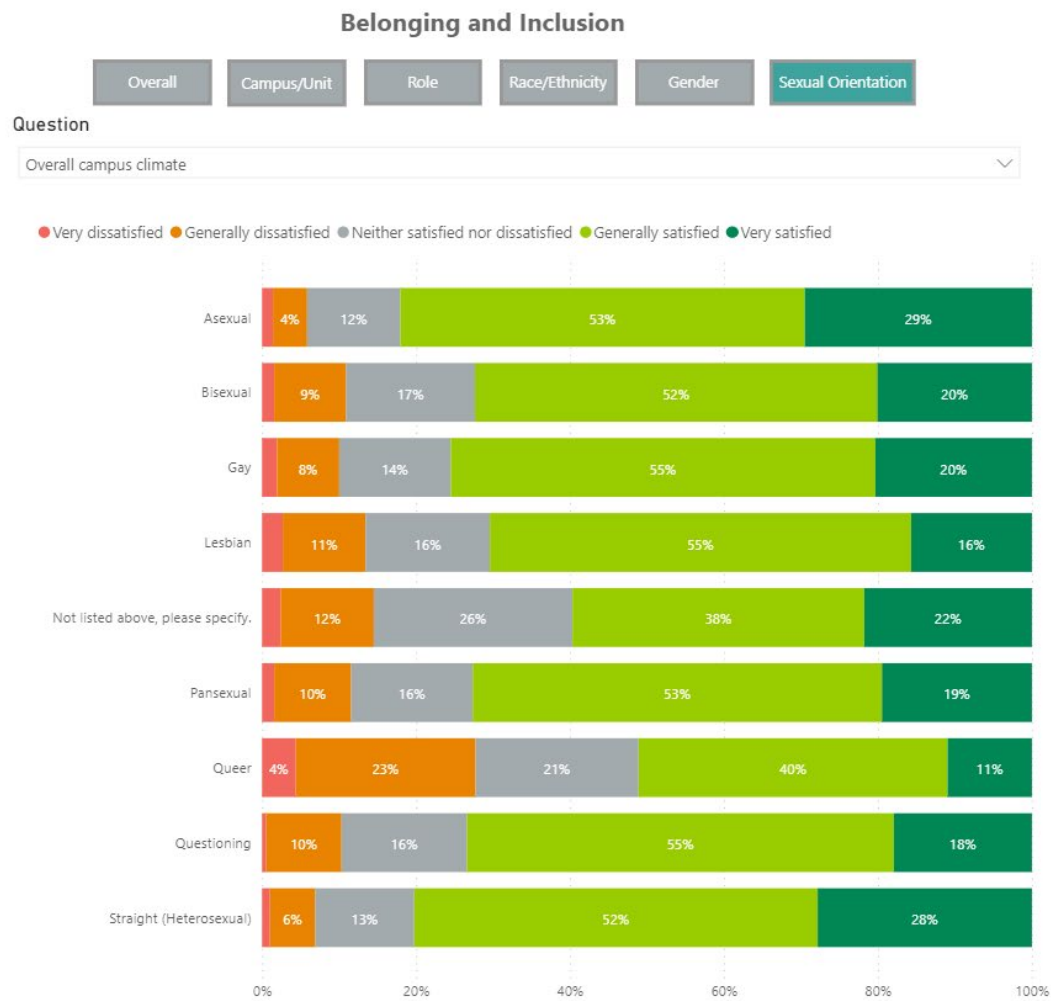
Sexual Orientation
All

Disability
All

Veteran
All

Political View
All

Religious Affiliation
All



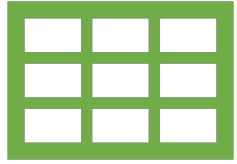
With a solid data model for your foundation, you are ready to start building

The image to the left is from Penn State's Community Survey dashboard, <https://opair.psu.edu/community-survey/dashboards/>

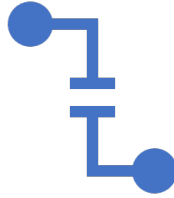
Just a sample of the many resources available to help you on your data visualization journey

- Microsoft Power BI
 - Guy in a Cube, https://www.youtube.com/watch?v=B_bAKrXAUh0
 - Microsoft Learning Paths, https://docs.microsoft.com/en-us/learn/powerplatform/power-bi?WT.mc_id=powerbi_landingpage-docs-link
 - Microsoft Power BI YouTube, <https://www.youtube.com/playlist?list=PL1N57mwBHtN0JFoKSR0n-tBkUJHeMP2cP&app=desktop>
 - Power BI Blog, <https://powerbi.microsoft.com/en-us/blog/>
- Tableau
 - Tableau Training videos, <https://www.tableau.com/learn/training/20202>
- LinkedIn Learning, Coursera, Udemy, Udacity, Edx

In recap, we hope that you can now:



Describe flat files
and their
limitations



Sketch out a
basic relational
data model



List the key steps
to build a data
model



Find additional
resources for
self-guided
learning

...so that you can get on with the fun part – visualizing your data!

Thank you!

- Betty Harper, bharper@psu.edu
- Shannon Lee, slee@psu.edu
- Carly Sunseri, cjt23@psu.edu