

2024
**Research
Analytics
Summit**

Building Capacity, Empowering Skill Transfer,
and Energizing a Growing Community

Data Models

Robert Pilgrim
rpilgrim@uark.edu

Check out my blog



Connect with me



The background features a dark blue and black color scheme with abstract data visualizations. A white line graph with circular markers is prominent in the upper left. To the right, there are vertical bars and grid lines, suggesting a bar chart or data table. The overall aesthetic is clean and modern, typical of a professional presentation.

Data Models

The Foundation of any Lean Team

Who am I



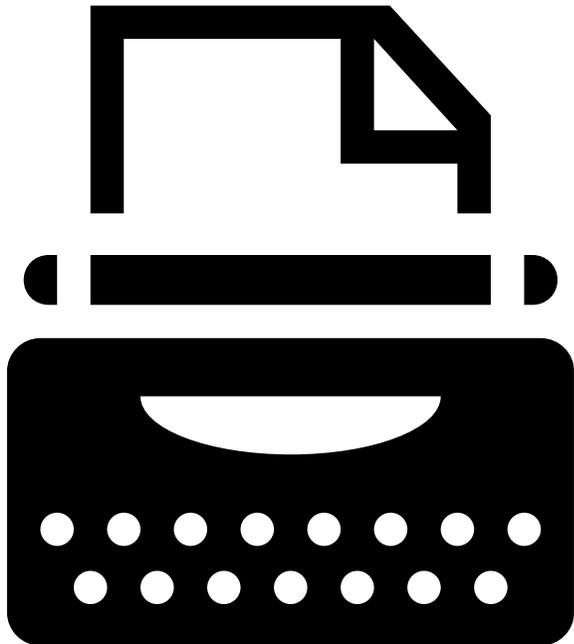
Robert Pilgrim, rpilgrim@uark.edu



Associate Director of Data Strategy & Insights

Division of Research & Innovation, University of Arkansas

Overview

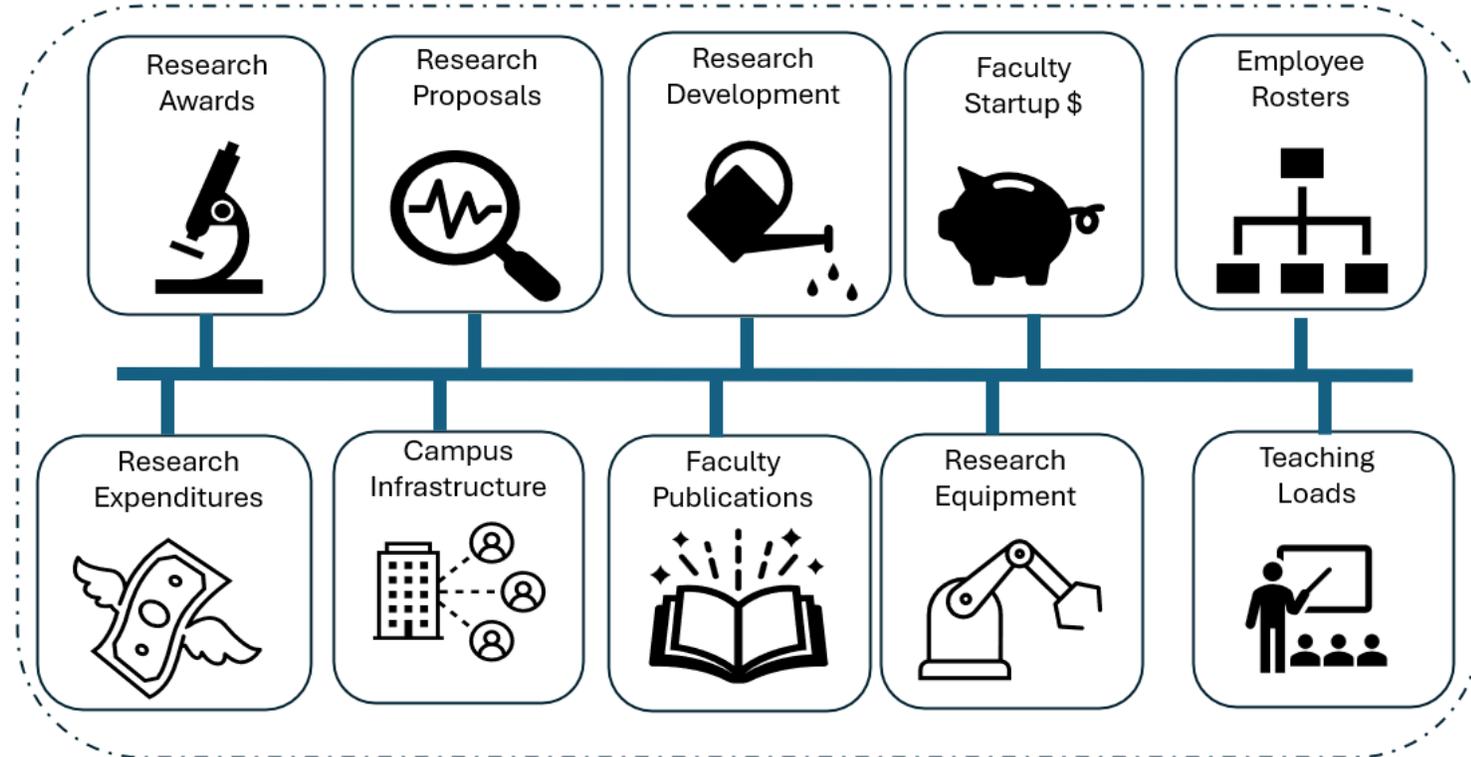


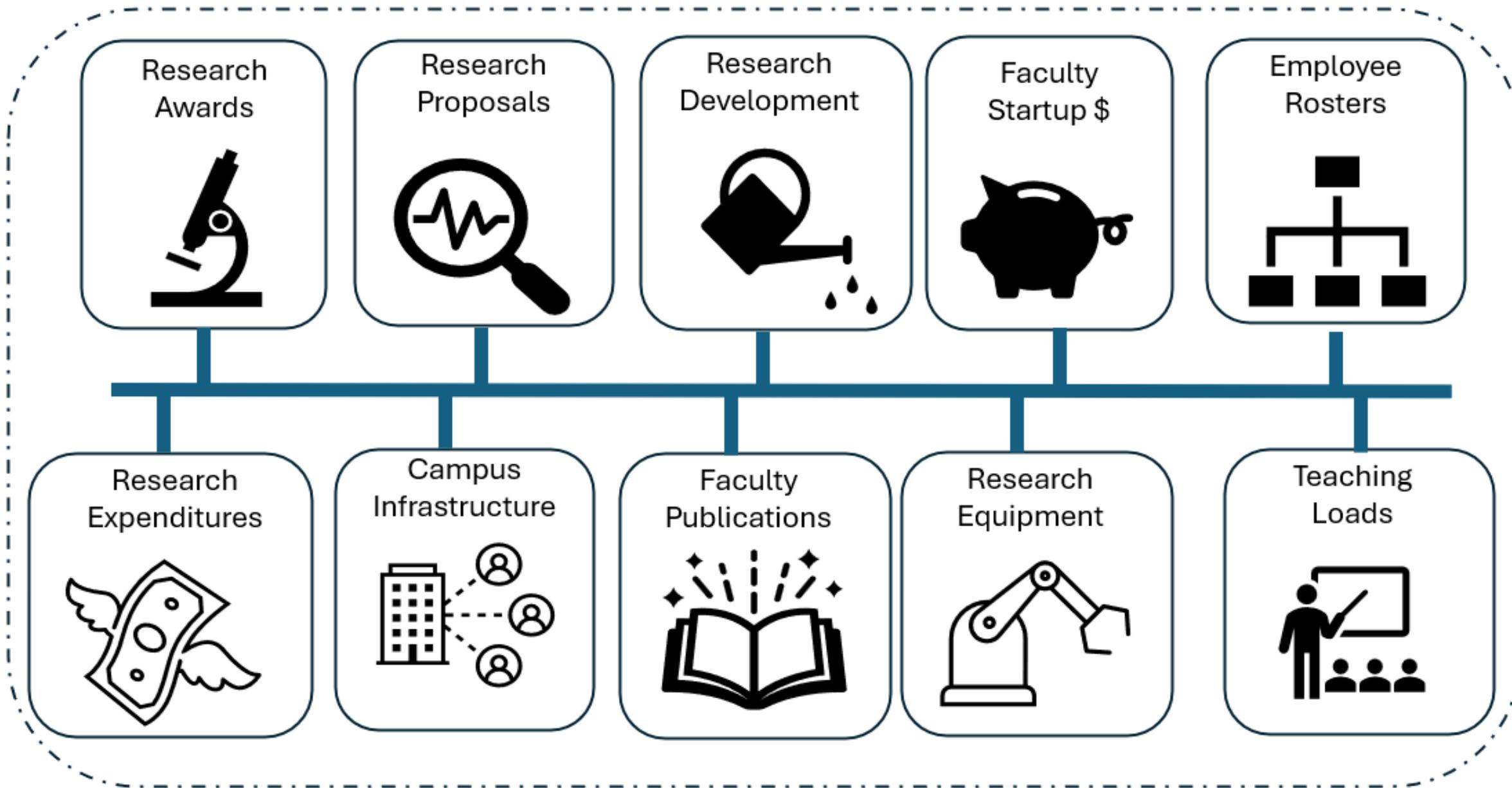
- What is a Data Model (DM)
- Benefits of Using a DM
- Focus on Sponsored Research related data
 - Campus Strategic Model
- Cross-Data calculations
 - DAX (Data Analysis Expressions) Language
- Pivot Tables
- Building with your Data Model

What is a Data Model?

- **Blueprint** of how data is “**Connected**”
- Easy **Cross-Data** source **calculations**
- Specialized **Functions**
- Data Models are accessed via **Pivot Tables**
- Multiple report creators

Close up next slide





What is Special about a Data Model

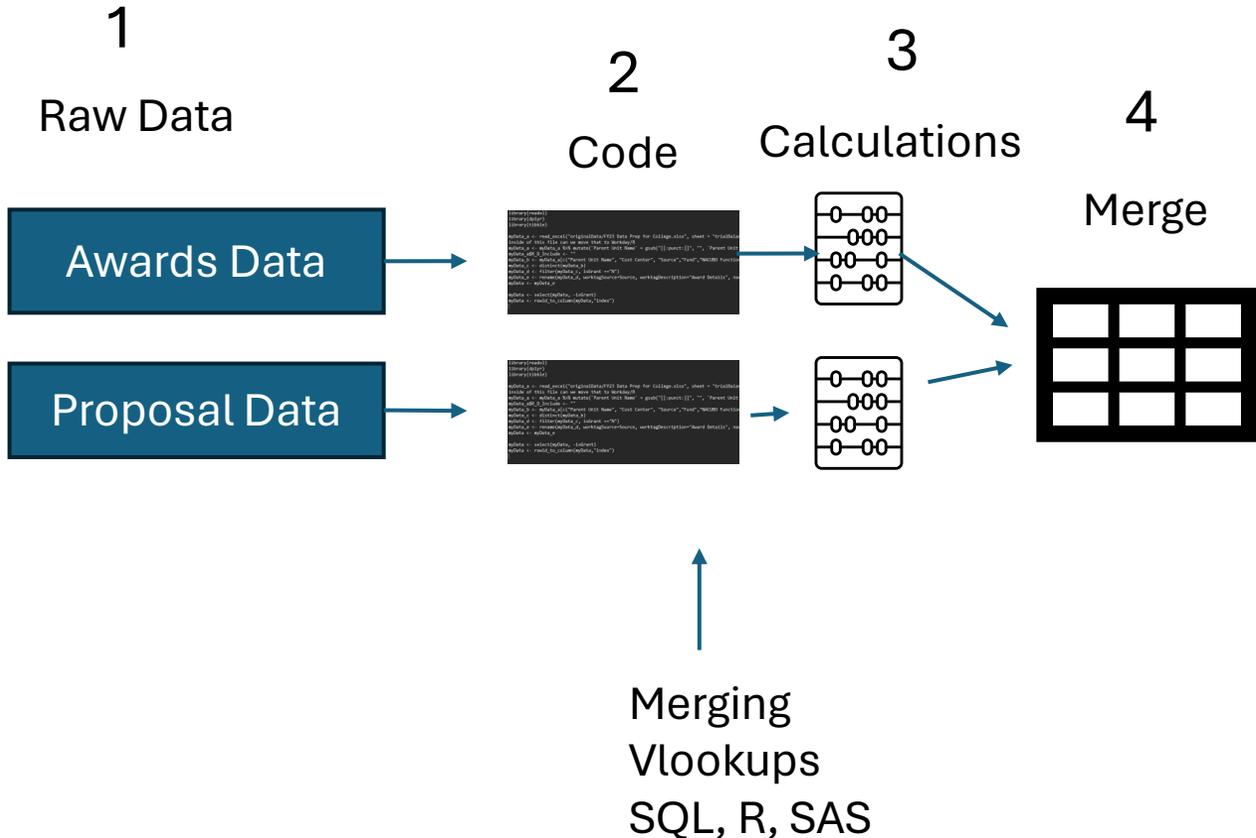
- Dedicated functions for **cross-data calculations**
- Single platform for start to finish reporting and analytics
 - Load & Transform data
 - Create and deploy Data Model
 - Handle security and permissions (including row level security)
 - Publish Reports
 - External embedding options

- Report creations via Pivot Tables

Typical “Problematic” Workflow



Request: We need by Fiscal Year:
Award \$, Proposal Success Rates, Proposal Counts, Total unique PI's



FY	Amount Award	Count Proposals PI	Count Proposals Funded	pct Funded
2024	\$249,724,815	652	212	18.7%
2023	\$447,368,567	965	788	36.2%
2022	\$437,849,800	946	953	44.6%
2021	\$317,595,905	933	839	38.0%
2020	\$254,922,705	919	852	38.9%
2019	\$276,673,478	848	748	38.9%
2018	\$204,752,204	824	681	35.8%
Grand Total	\$2,188,887,474	2,254	5,073	37.1%



Mission Complete

More Realistic Example

Initial Request | Campus

Award \$



Proposal Success Rates,

Proposal Counts,

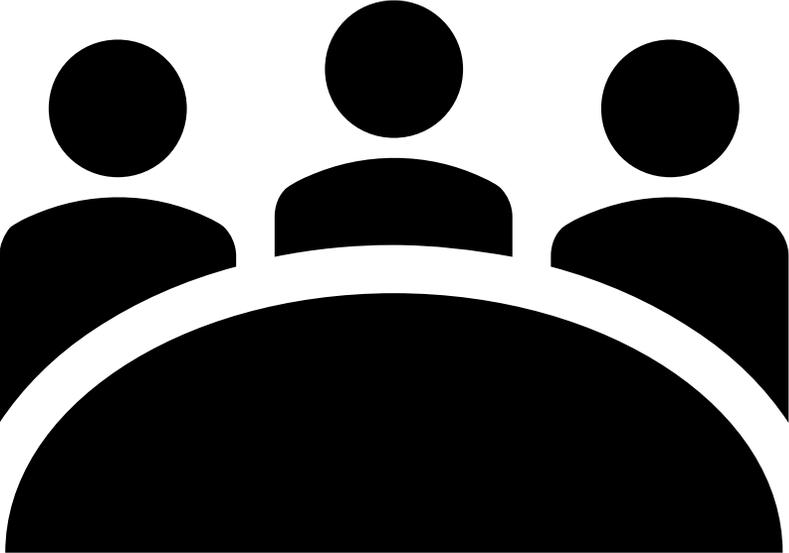
Total unique PI's

More Realistic Example

**Initial Request | Campus
Award \$**



**Proposal Success Rates,
Proposal Counts,
Total unique PI's**



More Realistic Example

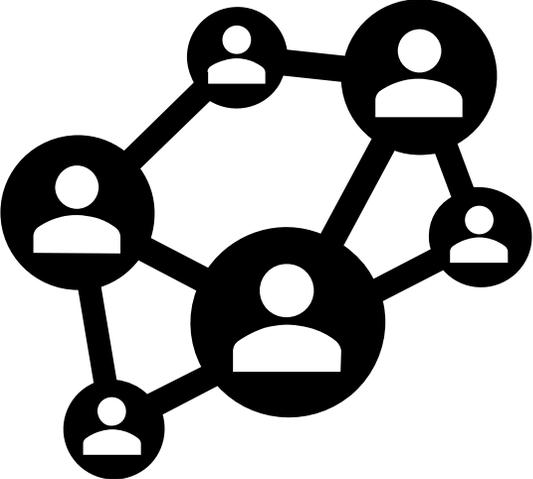
Initial Request | Campus
Award \$
Proposal Success Rates,
Proposal Counts,
Total unique PI's



By College

Can you send the merged data to X so they can build some reports

Remove People



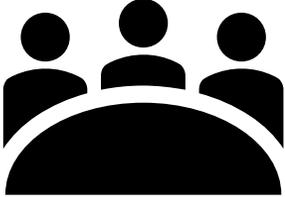
Exclude Sponsors

My calculations don't match yours?

More Realistic Example

Initial Request | Campus Award \$

**Proposal Success Rates,
Proposal Counts,
Total unique PI's**



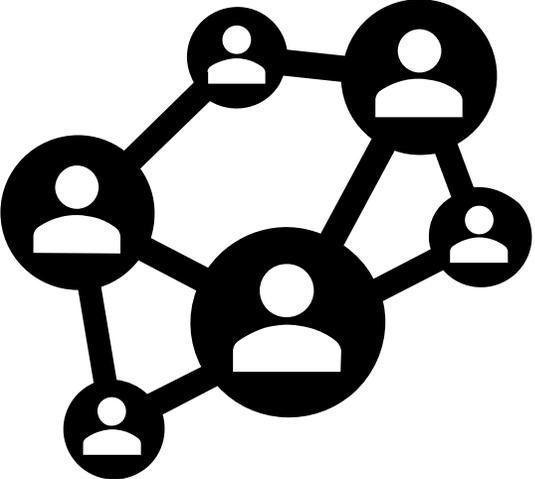
Faculty Level

We printed your report a month ago, but the numbers don't match, what changed?



Remove People

By College



Can you send the merged data to X so they can build some reports



How do Tenure Track vs Tenured faculty metrics compare?

Exclude Sponsors



Where is the most recent version of the data?

Can you put this into a line chart and do some trend analysis

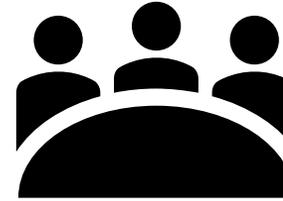
My calculations don't match yours?

Multiple Code

Initial Request | Campus Award \$



Proposal Success Rates, Proposal Counts, Total unique PI's



```
library(readr)
library(dplyr)
library(tibble)

myData_a <- read_excel("originalData/FY23 Data Prep for College.xlsx", sheet = "trialData")
# Inside of this file can we move that to Workbook?
myData_a <- myData_a %>% mutate(Parent Unit Name = gsub("[{}]", "", Parent Unit
myData_a$D_Include <- ""
myData_b <- myData_a[!(Parent Unit Name ~ "Cost Center", "Source", "Fund", "MACRO Function
myData_c <- distinct(myData_b)
myData_d <- filter(myData_c, isGrant == "N")
myData_e <- rename(myData_d, worktagSource=Source, worktagDescription="Award Details", na
myData <- myData_e

myData <- select(myData, -isGrant)
myData <- read_to_column(myData, "index")
```

Faculty Level

```
library(readr)
library(dplyr)
library(tibble)

myData_a <- read_excel("originalData/FY23 Data Prep for College.xlsx", sheet = "trialData")
# Inside of this file can we move that to Workbook?
myData_a <- myData_a %>% mutate(Parent Unit Name = gsub("[{}]", "", Parent Unit
myData_a$D_Include <- ""
myData_b <- myData_a[!(Parent Unit Name ~ "Cost Center", "Source", "Fund", "MACRO Function
myData_c <- distinct(myData_b)
myData_d <- filter(myData_c, isGrant == "N")
myData_e <- rename(myData_d, worktagSource=Source, worktagDescription="Award Details", na
myData <- myData_e

myData <- select(myData, -isGrant)
myData <- read_to_column(myData, "index")
```

We printed your report a months ago but the numbers don't match, what changed?

```
library(readr)
library(dplyr)
library(tibble)

myData_a <- read_excel("originalData/FY23 Data Prep for College.xlsx", sheet = "trialData")
# Inside of this file can we move that to Workbook?
myData_a <- myData_a %>% mutate(Parent Unit Name = gsub("[{}]", "", Parent Unit
myData_a$D_Include <- ""
myData_b <- myData_a[!(Parent Unit Name ~ "Cost Center", "Source", "Fund", "MACRO Function
myData_c <- distinct(myData_b)
myData_d <- filter(myData_c, isGrant == "N")
myData_e <- rename(myData_d, worktagSource=Source, worktagDescription="Award Details", na
myData <- myData_e

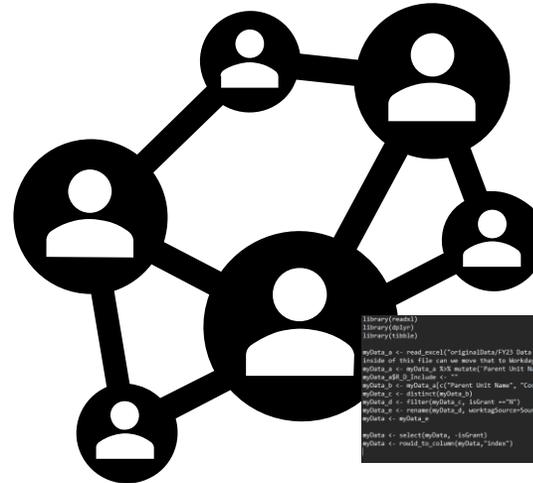
myData <- select(myData, -isGrant)
myData <- read_to_column(myData, "index")
```

By College

Can you send the merged data to X so they can build some reports



Remove People



How do Tenure Track vs Tenured faculty metrics compare?

Where is the most recent version of the data?

```
library(readr)
library(dplyr)
library(tibble)

myData_a <- read_excel("originalData/FY23 Data Prep for College.xlsx", sheet = "trialData")
# Inside of this file can we move that to Workbook?
myData_a <- myData_a %>% mutate(Parent Unit Name = gsub("[{}]", "", Parent Unit
myData_a$D_Include <- ""
myData_b <- myData_a[!(Parent Unit Name ~ "Cost Center", "Source", "Fund", "MACRO Function
myData_c <- distinct(myData_b)
myData_d <- filter(myData_c, isGrant == "N")
myData_e <- rename(myData_d, worktagSource=Source, worktagDescription="Award Details", na
myData <- myData_e

myData <- select(myData, -isGrant)
myData <- read_to_column(myData, "index")
```

Can you put this into a line chart and do some trend analysis

My calculations don't match yours?

Exclude Sponsors



```
library(readr)
library(dplyr)
library(tibble)

myData_a <- read_excel("originalData/FY23 Data Prep for College.xlsx", sheet = "trialData")
# Inside of this file can we move that to Workbook?
myData_a <- myData_a %>% mutate(Parent Unit Name = gsub("[{}]", "", Parent Unit
myData_a$D_Include <- ""
myData_b <- myData_a[!(Parent Unit Name ~ "Cost Center", "Source", "Fund", "MACRO Function
myData_c <- distinct(myData_b)
myData_d <- filter(myData_c, isGrant == "N")
myData_e <- rename(myData_d, worktagSource=Source, worktagDescription="Award Details", na
myData <- myData_e

myData <- select(myData, -isGrant)
myData <- read_to_column(myData, "index")
```

Where is Most of the Time Spent ?

Initial Request
Award \$
Proposal Success Rates,
Proposal Counts,
Total unique PI's



Making the Metrics

1% Effort
(Once initial calculations
created)

Making the Reports

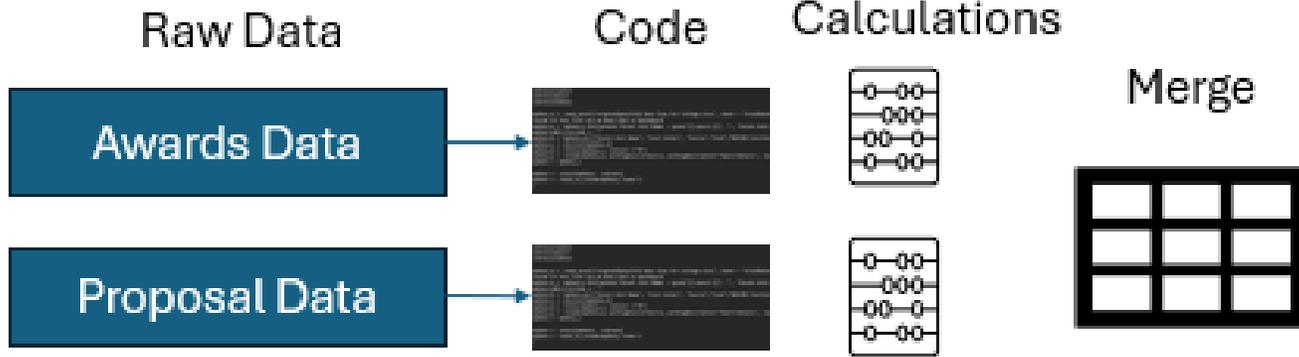
99% Effort

- We printed your report a months ago but the numbers don't match, what changed?
- We need this at the faculty level
- College Y have made some calculations that do not match yours, can you help them
- Make this for each College
- ENGR want these people removed as they left
- Can you send the merged data to X so they can build some reports
- How do Tenure Track vs Tenured faculty metrics compare?
- Where is the most recent version of the data?
- Don't include this expense category for unit X and Y
- Can you put this into a line chart and do some trend analysis

Workflow Paths

Data Model

Reporting



FY	Amount Award	Count	Count	pct
		Proposals	Proposals	
	PI	Funded	Funded	
2024	\$249,724,815	652	212	18.7%
2023	\$447,368,567	965	788	36.2%
2022	\$437,849,800	946	953	44.6%
2021	\$317,595,905	933	839	38.0%
2020	\$254,922,705	919	852	38.9%
2019	\$276,673,478	848	748	38.9%
2018	\$204,752,204	824	681	35.8%
Grand Total	\$2,188,887,474	2,254	5,073	37.1%

Separate Reporting from Data Model

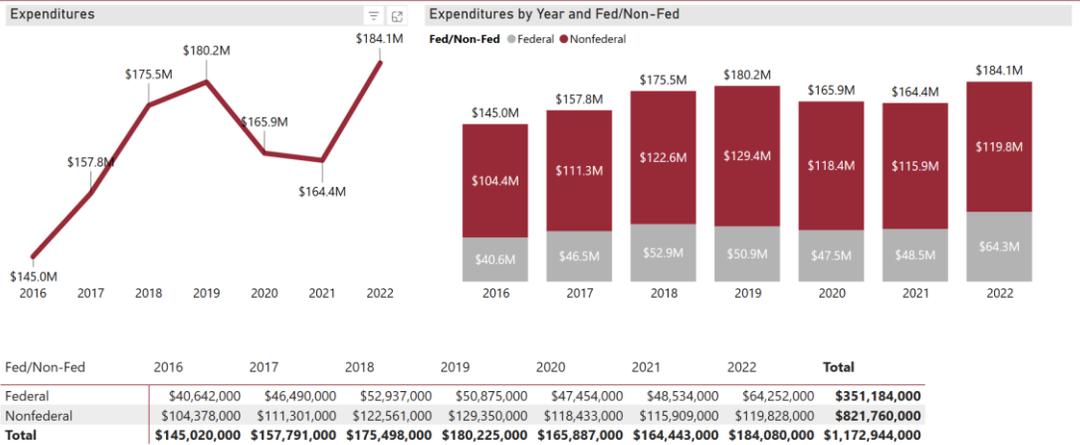
Data Model

- Design
- Create



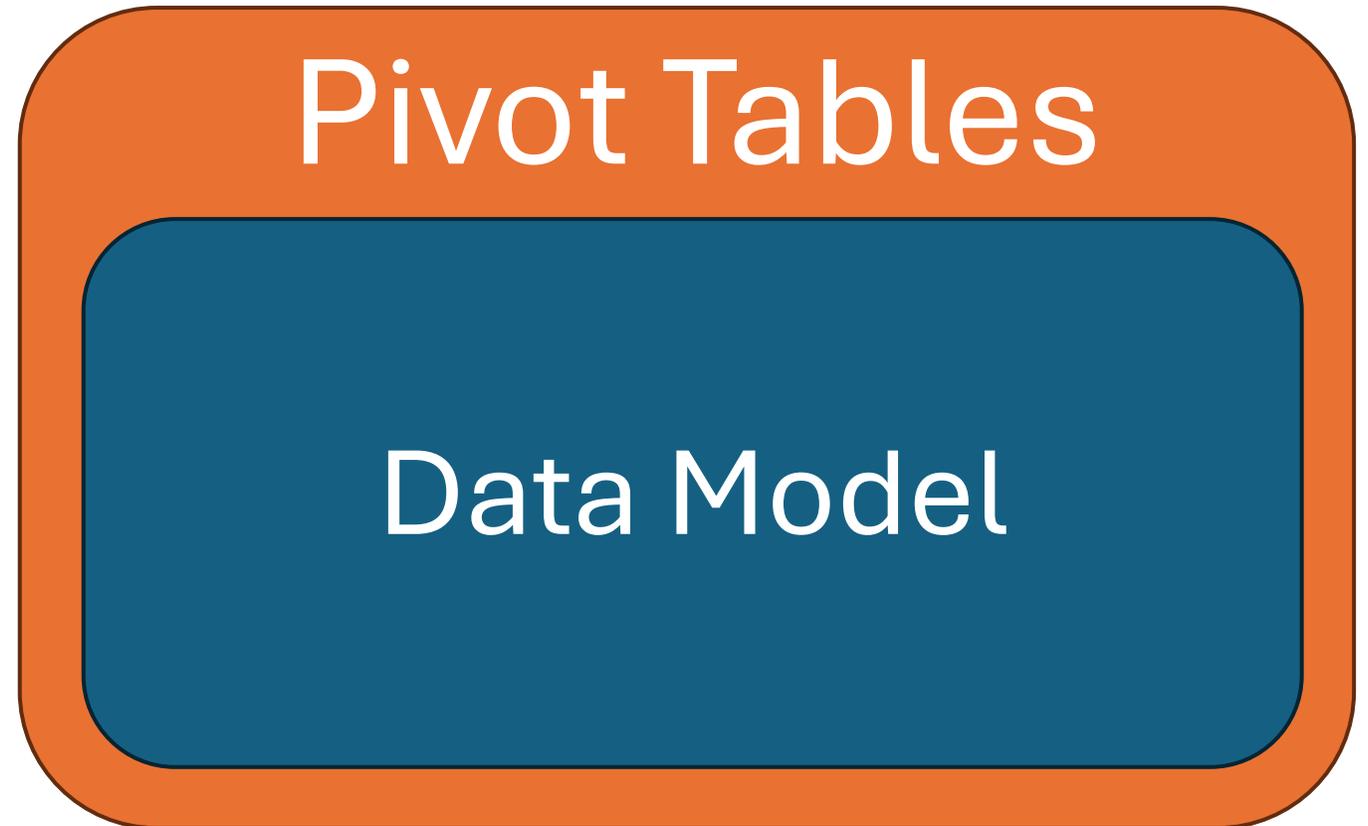
Reporting

- Create
- Distribute



Pivot Tables are the Foundation of Data Models

- Data Models are “Accessed” by Pivot Tables





Sponsored Research Data Models

Awards and Proposals

Data Models=Connections NOT Merging

Research Awards

Award **Number**

Fiscal Year

Award **Title**

Award **PI**

Award **Investigator**

Award **Investigator Role**

Award **Date**

Award **Investigator College**

Award **Sponsor**

Award **Sponsor Type**

Award **Sponsor Type Group**

Award **Amount**

Research Proposals

Proposal **Number**

Fiscal Year

Proposal **Title**

Proposal **PI**

Proposal **Investigator**

Proposal **Investigator Role**

Proposal **Date**

Proposal **Investigator College**

Proposal **Sponsor**

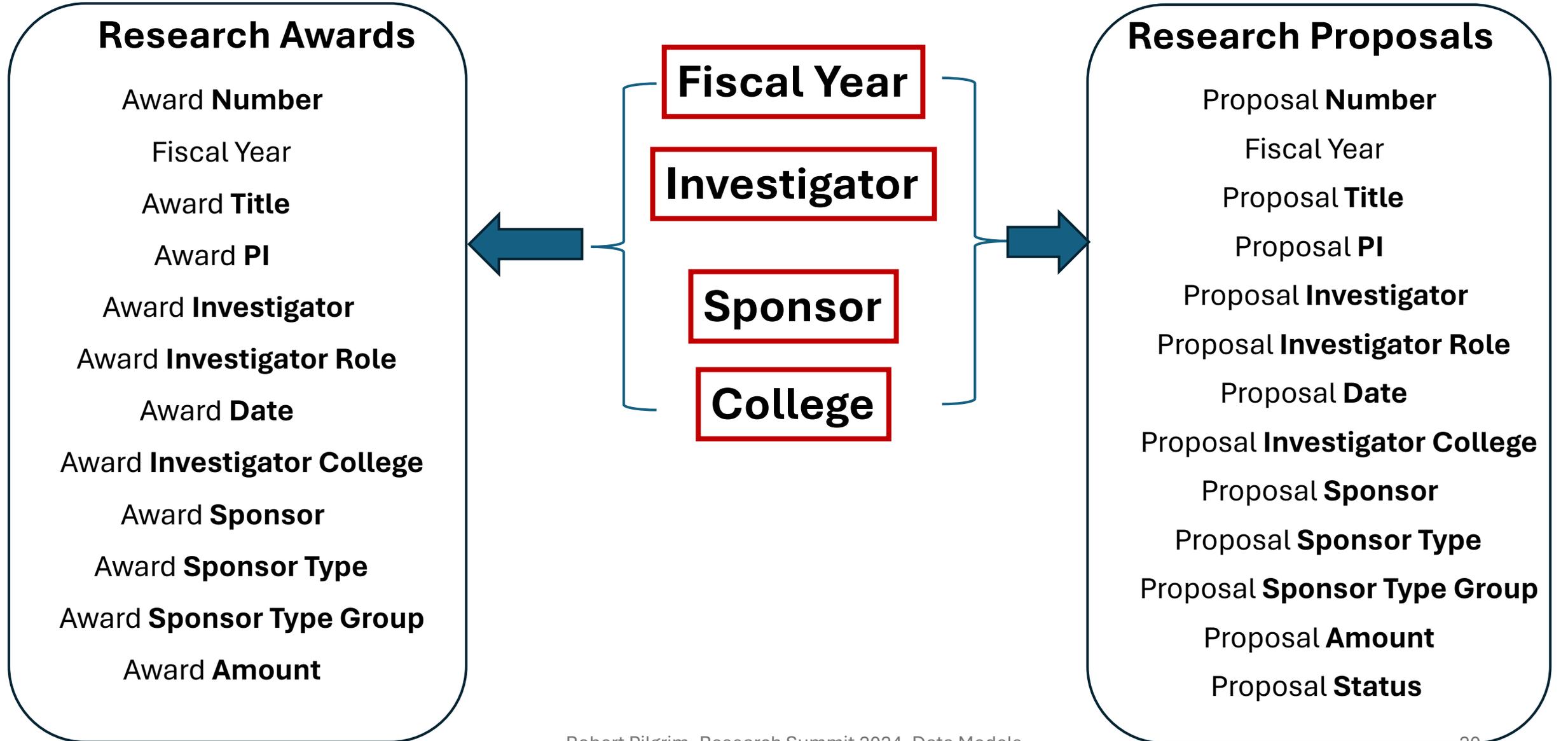
Proposal **Sponsor Type**

Proposal **Sponsor Type Group**

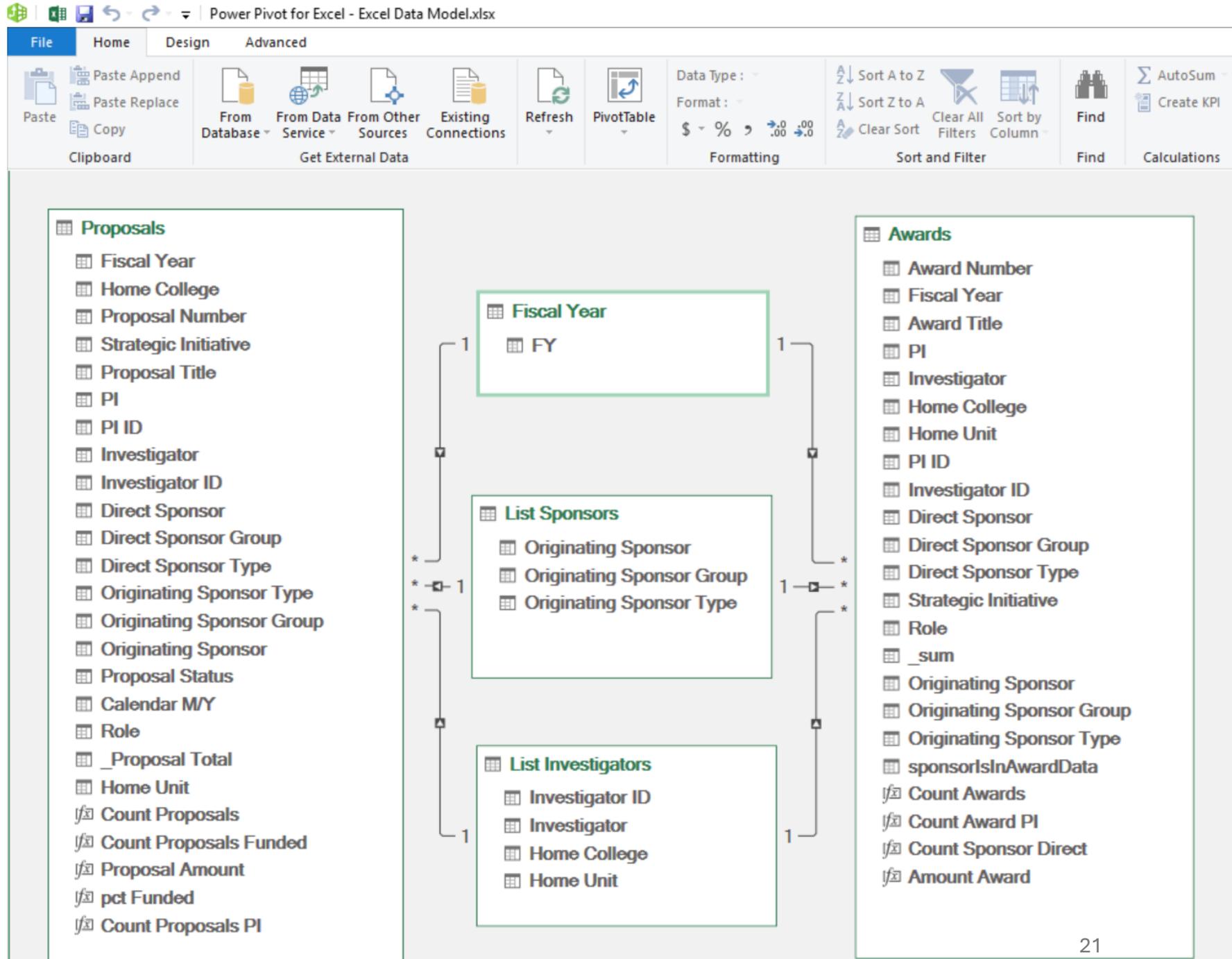
Proposal **Amount**

Proposal **Status**

“Connect” Data



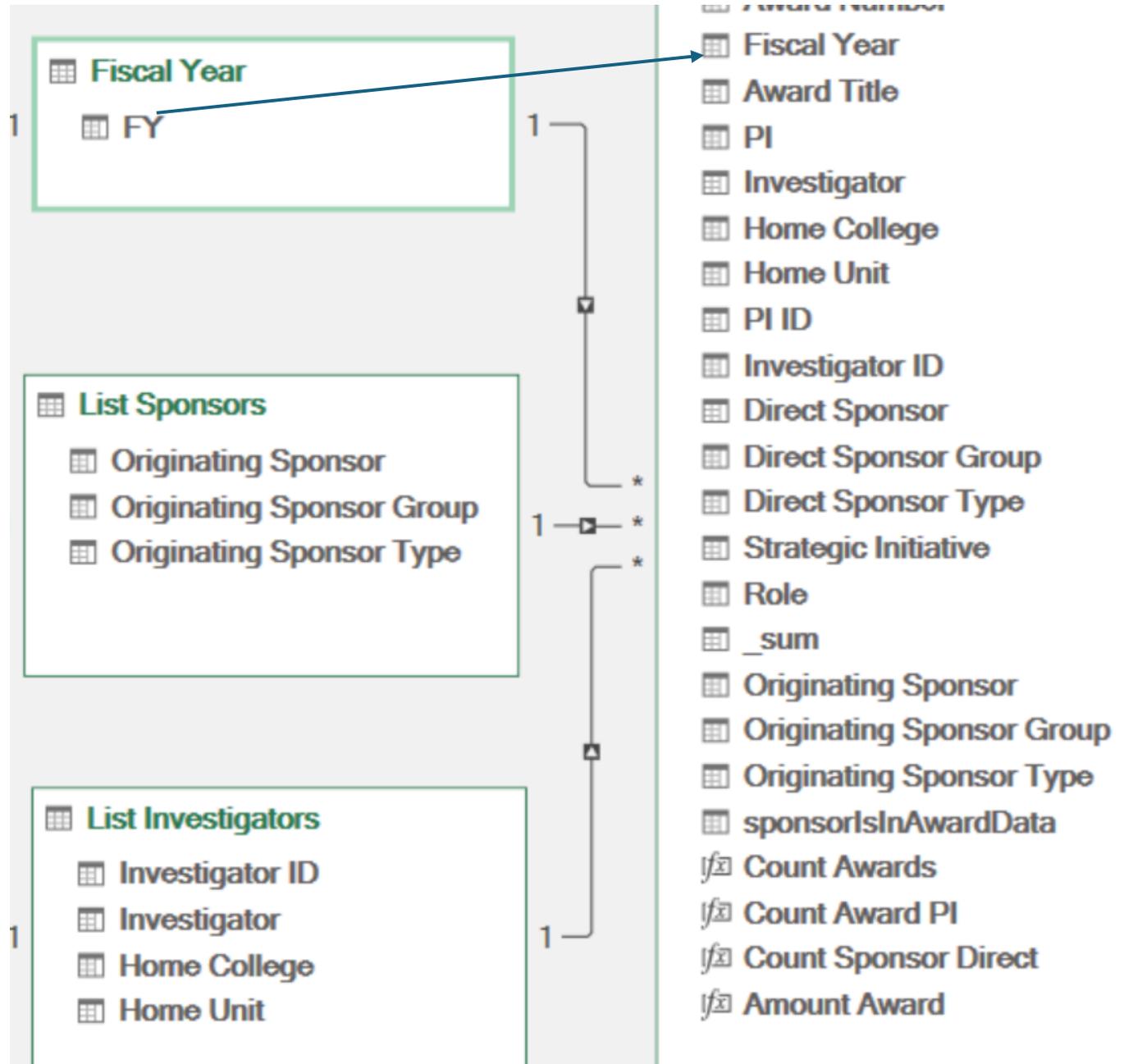
Excel Data Model

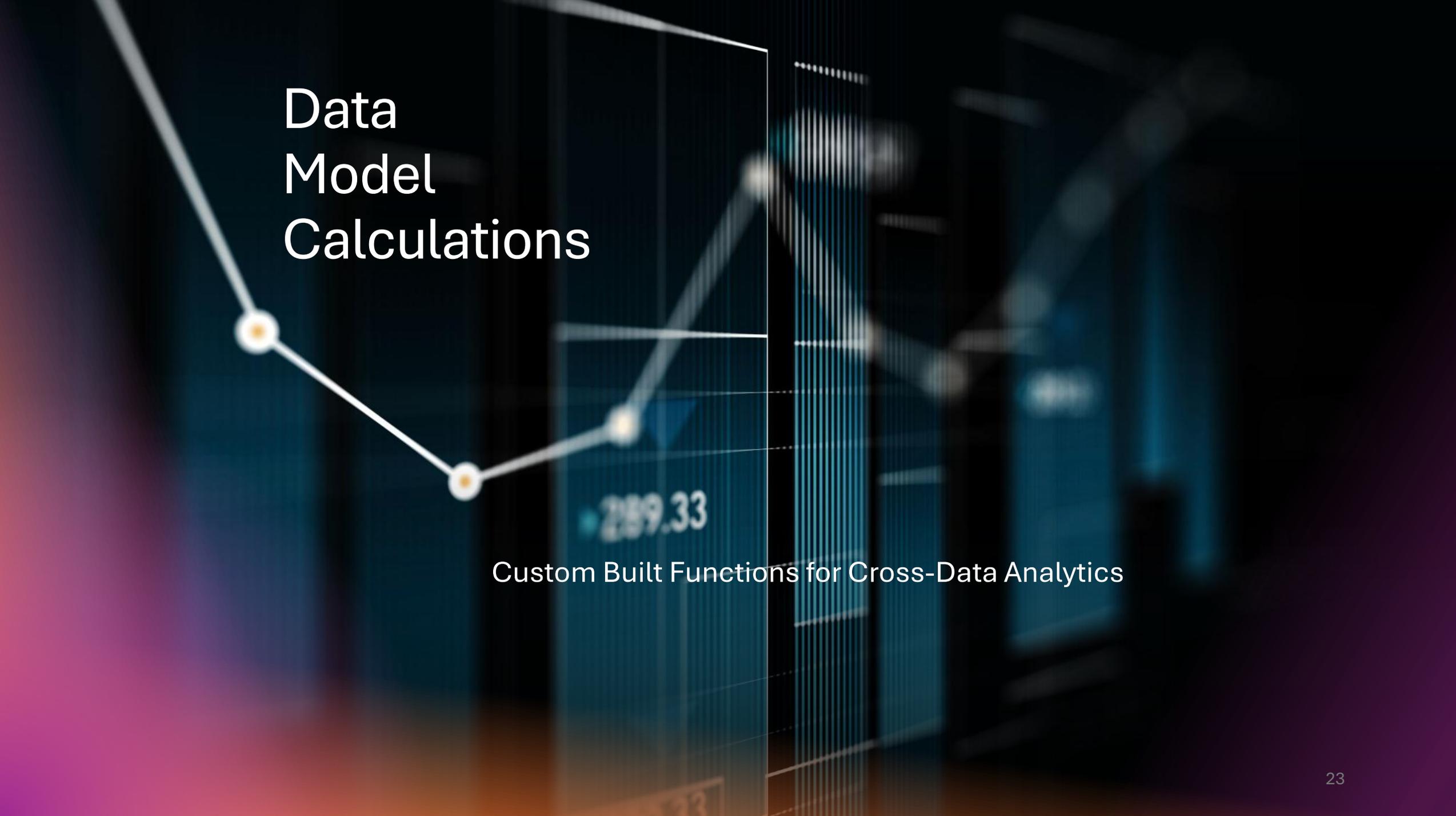


Excel Data Model



Drag Connecting Variables onto Matching variables in





Data Model Calculations

Custom Built Functions for Cross-Data Analytics

Hello World | DAX Language

Data Analysis Expressions (DAX)

Calculation
Name

Table

```
Count Awards:=DISTINCTCOUNT(Awards[Award Number])
```

Function

Variable

A screenshot of a data model table list. The table 'Awards' is expanded, showing its columns. The column 'Award Number' is highlighted with a red rectangular box. A blue arrow points from this box to the 'Award Number' variable in the DAX formula shown in the diagram.

Awards	
Award Number	
Fiscal Year	
Award Title	
PI	
Investigator	
Home College	
Home Unit	
PI ID	
Investigator ID	
Direct Sponsor	
Direct Sponsor Group	
Direct Sponsor Type	
Strategic Initiative	
Role	
_sum	
Originating Sponsor	
Originating Sponsor Group	
Originating Sponsor Type	
sponsorIsInAwardData	
Count Awards	
Count Award PI	
Count Sponsor Direct	
Amount Award	

Base DAX Calculations

Research
Awards

Count Awards:=**DISTINCTCOUNT**(Awards[Award Number])
Amount Award:=**SUM**([_sum])

Research
Proposals

Count PI=**DISTINCTCOUNT**(Proposals[PI])

Count Proposals:=**DISTINCTCOUNT**(Proposals[Proposal Number])

Count Proposals Funded:=

CALCULATE([**Count Proposals**],Proposals[Proposal Status]="Funded")

Modifier

Filter for "Funded" then do [Count Proposals]

Campus
Infrastructure

Research Area:=**SUM**([sqft])

Research
Expenditures

Research Exp.:=**SUM**([exp])

Cros-Data Calculations

- [Count Awards]
- [Count Proposals]
- [Count PI]
- [Count Proposals Funded]
- [Research Area]
- [Research Exp.]

1. Exp per sqft of Research Space

[Research Exp.] / [Research Area]

2. Proposal Success Rate

[Count Proposals Funded] / [Count Proposals]

3. Avg Exp per PI

[Research Exp.] / [Count PI]

4. Avg Proposal Submission Rate per PI

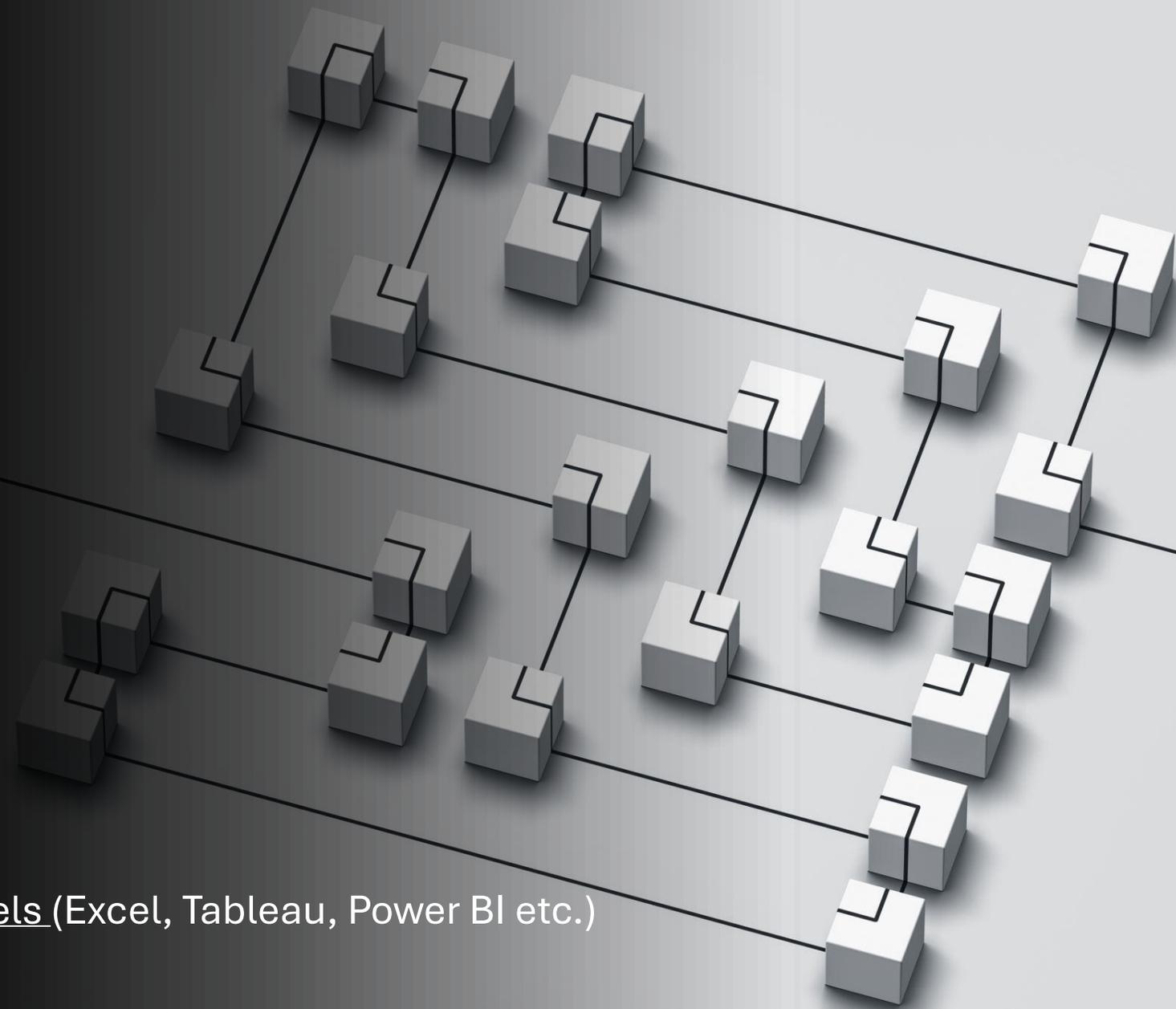
[Count Proposals] / [Count PI]

Bringing this all Together

Data Model + Pivot Table



The Concept of a Pivot Table



Foundation Element of All Data Models (Excel, Tableau, Power BI etc.)

Pivot Tables 101

Row Labels	Sum of _sum
2024	\$250,351,073
2023	\$447,608,567
2022	\$438,148,800
2021	\$317,595,905
2020	\$255,016,878
2019	\$276,841,837
2018	\$205,866,572
Grand Total	\$2,191,429,632

Aggregation Calculations e.g. SUM

add to report:

Search

- Award Number
- Fiscal Year**
- Award Title
- PI
- Investigator
- Home College
- Home Unit
- PI ID
- Investigator ID
- Direct Sponsor
- Direct Sponsor Group
- Direct Sponsor Type
- Strategic Initiative
- Role
- _sum**
- Originating Sponsor
- Originating Sponsor G...
- Originating Sponsor T...
- sponsorIsInAwardData

More Tables...

Filters

Columns

Rows

Fiscal Year

Σ Values

Sum of _sum

29

Another Example

Amount	FY						
Row Labels	2018	2019	2020	2021	2022	2023	2024
Federal	\$129,747,247	\$168,849,582	\$168,628,495	\$185,016,861	\$240,779,271	\$234,973,857	\$187,803,835
Industry	\$9,784,598	\$14,034,896	\$10,786,063	\$17,705,443	\$15,234,564	\$13,134,975	\$12,038,421
Local	\$408,227	\$2,452,851	\$893,071	\$2,226,083	\$1,247,705	\$951,527	\$1,124,924
Nonprofit	\$31,507,772	\$34,066,601	\$30,065,738	\$47,254,318	\$72,301,634	\$82,726,153	\$33,528,045
Other	\$2,779,888	\$15,547,529	\$3,322,385	\$1,545,615	\$2,095,160	\$1,357,304	\$553,136
State	\$30,524,472	\$41,722,019	\$41,226,953	\$63,847,585	\$106,191,466	\$114,224,751	\$14,676,454
Grand Total	\$204,752,204	\$276,673,478	\$254,922,705	\$317,595,905	\$437,849,800	\$447,368,567	\$249,724,815

Choose fields to add to report:

Search

- Award Number
- Fiscal Year**
- Award Title
- PI
- Investigator
- Home College
- Home Unit
- PI ID
- Investigator ID
- Direct Sponsor
- Direct Sponsor Gr...
- Direct Sponsor Ty...
- Strategic Initiative
- Role
- _sum**
- Originating Spons...
- Originating Spon...**
- Originating Spons...
- sponsorIsInAward...

More Tables...

Filters

Columns

Fiscal Year

Rows

Originating Sponsor Group

Values

Amount

Behind the Scene Calculations

Amount	FY							
Row Labels		2018	2019	2020	2021	2022	2023	2024
Federal		\$129,747,247	\$168,849,582	\$168,628,495	\$185,016,861	\$240,779,271	\$234,973,857	\$187,803,835
Industry		\$9,784,598	\$14,034,896	\$10,786,063	\$17,705,443	\$15,234,564	\$13,134,975	\$12,038,421
Local		\$408,227	\$2,452,851	\$893,071	\$2,226,083	\$1,247,705	\$951,527	\$1,124,924
Nonprofit		\$31,507,772	\$34,066,601	\$30,065,738	\$47,254,318	\$72,301,634	\$82,726,153	\$33,528,045
Other		\$2,779,888	\$15,547,529	\$3,322,385	\$1,545,615	\$2,095,160	\$1,357,304	\$553,136
State		\$30,524,472	\$41,722,019	\$41,226,953	\$63,847,585	\$106,191,466	\$114,224,751	\$14,676,454
Grand Total		\$204,752,204	\$276,673,478	\$254,922,705	\$317,595,905	\$437,849,800	\$447,368,567	\$249,724,815

FY=2023 ***AND*** Sponsor Group="Nonprofit"

Drum Roll



Final Data Model | Build Some Reports

Proposals

Year
College
Unit
Emp ID
[Count Proposals]
[Count Proposals Funded]
[Count PI]

Awards

Year
College
Unit
Emp ID
[Count Awards]

Year

Unit

College

Emp ID

Cross Data Calculations

- Proposal Success Rate
- Exp per sqft of Research Space
- Avg Exp per PI
- Avg Proposal Submission Rate per PI

Research Expenditures

Year
College
Unit
Emp ID
[Research Exp.]

Research Space

Year
College
Unit
Emp ID
[Research Area]

Year

Unit

College

Emp ID

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Q1: Count Award and Count Proposals by College for Fiscal Year 2022

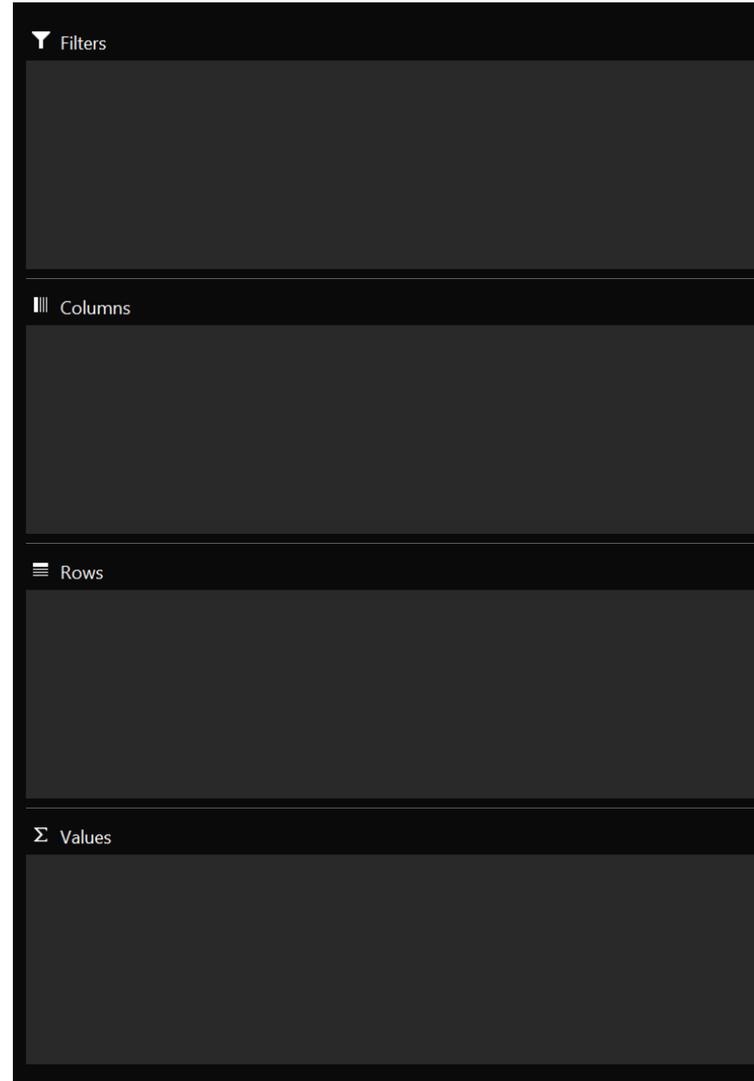
Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI



Unit

Emp ID

Q1: Count Award and Count Proposals by College for Fiscal Year 2022

[Count Proposals Funded]

[Count PI]

[Research Exp.]

[Research Area]

Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

The screenshot shows a dark-themed interface for a data visualization tool. It has four main sections: 'Filters' at the top with a dropdown arrow, containing a box labeled 'Year'; 'Columns' with a list icon, which is currently empty; 'Rows' with a list icon, containing a box labeled 'College'; and 'Σ Values' at the bottom, containing two blue text items: '[Count Awards]' and '[Count Proposals]'. The 'Year' and 'College' boxes are highlighted with green borders.

College

Unit

Year

Emp ID

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Q2: Count Award and Count Proposals by College for **ALL Fiscal Years**

Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

The screenshot shows a dark-themed interface for configuring a data visualization. It has four main sections: 'Filters' (with a dropdown arrow), 'Columns' (with a list icon), 'Rows' (with a list icon), and 'Values' (with a summation symbol). Each section is currently empty.

Unit

Emp ID

Q2: Count Award and Count Proposals by College for ALL Fiscal Years

[Count Proposals Funded]

[Count PI]

[Research Exp.]

[Research Area]

Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

Filters

Columns

Year

Rows

College

Σ Values

[Count Awards]
[Count Proposals]

College

Unit

Year

Emp ID

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Q3: avg Proposal Submission Rate for ENGR Units in 2022
AND Exp per sqft Research Space

Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

The screenshot shows a dark-themed interface for configuring a data visualization. It is divided into four main sections: 'Filters' (with a dropdown arrow), 'Columns' (with a list icon), 'Rows' (with a list icon), and 'Values' (with a summation symbol). Each section is currently empty, indicating that no filters, columns, rows, or values have been selected or defined yet.

Emp ID

Q3: avg Proposal Submission Rate for ENGR Units in 2022
AND Exp per sqft Research Space

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Cross Data Calculations

Proposal Success Rate

Avg Exp per PI

The screenshot shows a data visualization tool interface with the following sections:

- Filters:** Contains two filter boxes labeled "Year" and "College", both highlighted with green borders.
- Columns:** An empty section for defining column-level calculations.
- Rows:** Contains one row box labeled "Unit", highlighted with a green border.
- Σ Values:** A section for defining aggregate values, containing two boxes:
 - A box labeled **Avg Proposal Submission Rate per PI**, highlighted with an orange border.
 - A box labeled **Exp per sqft of Research Space**, highlighted with an orange border.

College

Year

Unit

Emp ID

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Q4: By PI in ENGR avg Proposal Submission Rate
AND Exp per sqft Research Space, all years

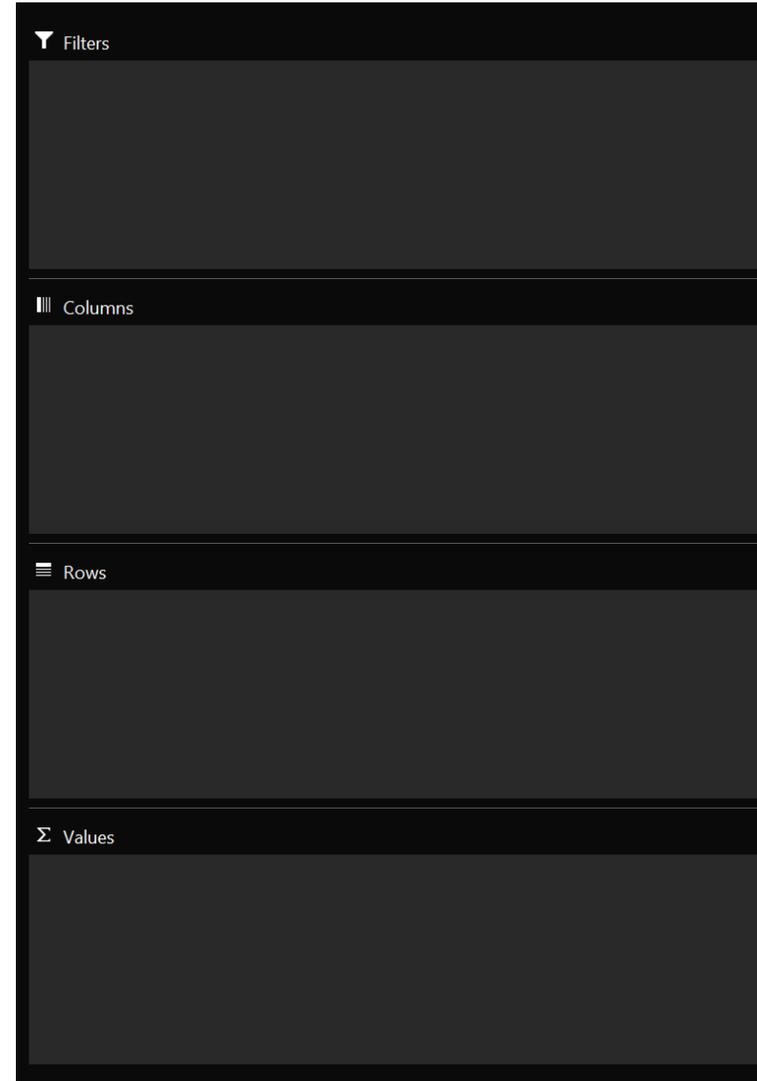
Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI



Unit

Q4: By PI in ENGR avg Proposal Submission Rate
AND Exp per sqft Research Space, all years

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Cross Data Calculations

Proposal Success Rate

Avg Exp per PI

The screenshot shows a data visualization tool interface with the following sections:

- Filters:** A dropdown menu is open, showing the word "College" highlighted with a green border.
- Columns:** A list of columns is shown, with "Year" highlighted by a green border.
- Rows:** A list of rows is shown, with "Emp ID" highlighted by a green border.
- Σ Values:** A section for calculated values, containing two items:
 - Avg Proposal Submission Rate per PI** (highlighted with an orange border)
 - Exp per sqft of Research Space** (highlighted with an orange border)

All Metrics by Unit and Year

Aggregation

Row

Unit

[Count Proposals]
[Count Proposals Funded]
[Count PI]
[Count Awards]
[Research Exp.]
[Research Area]

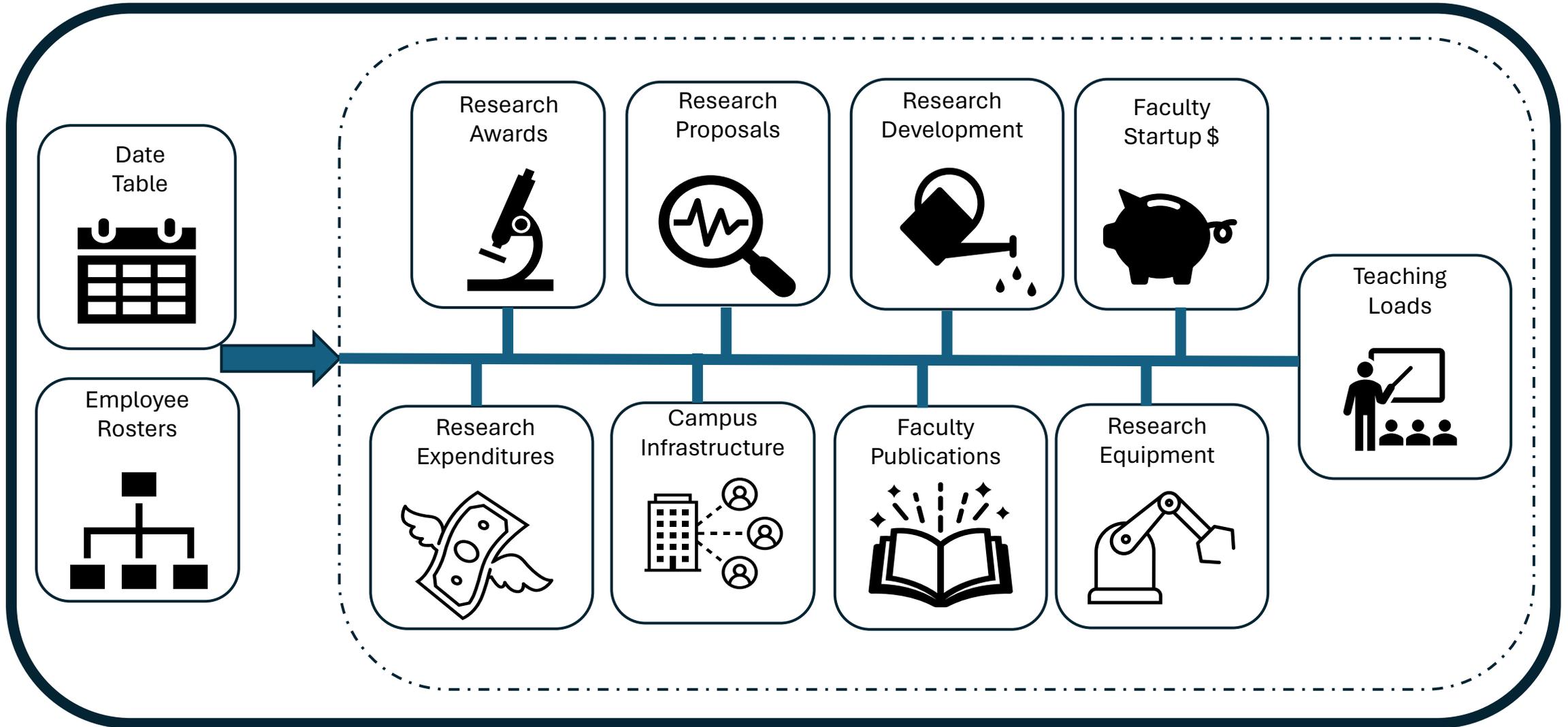
Proposal Success Rate

Avg Exp per PI

Avg Proposal Submission Rate per PI

Exp per sqft of Research Space

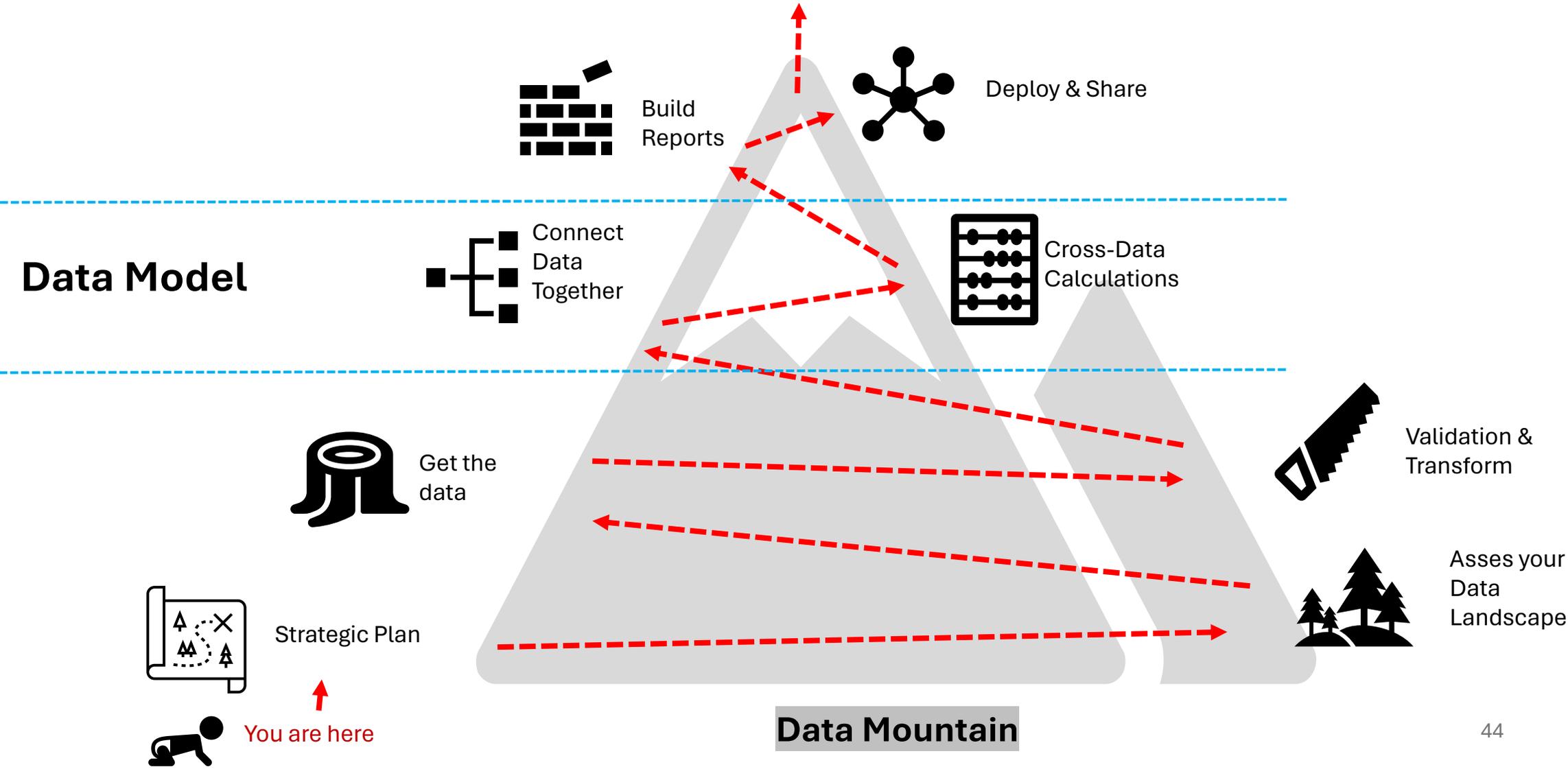
Typical University Data Model (Research Focused)



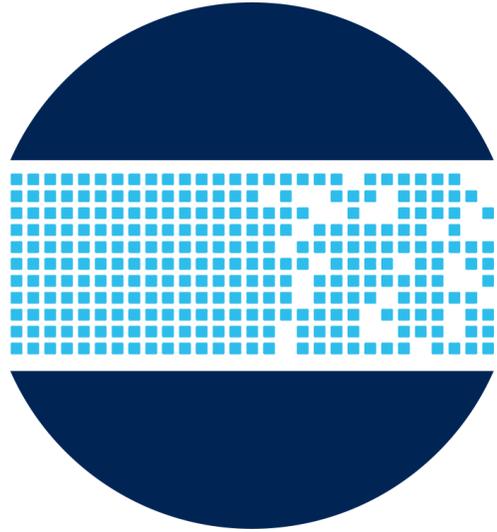
Your Analytics Journey



Get that promotion and live happily ever after



Data Mountain



2024
**Research
Analytics
Summit**
Building Capacity, Empowering Skill Transfer,
and Energizing a Growing Community

Data Models

Robert Pilgrim
rpilgrim@uark.edu

Check out my blog



Connect with me

