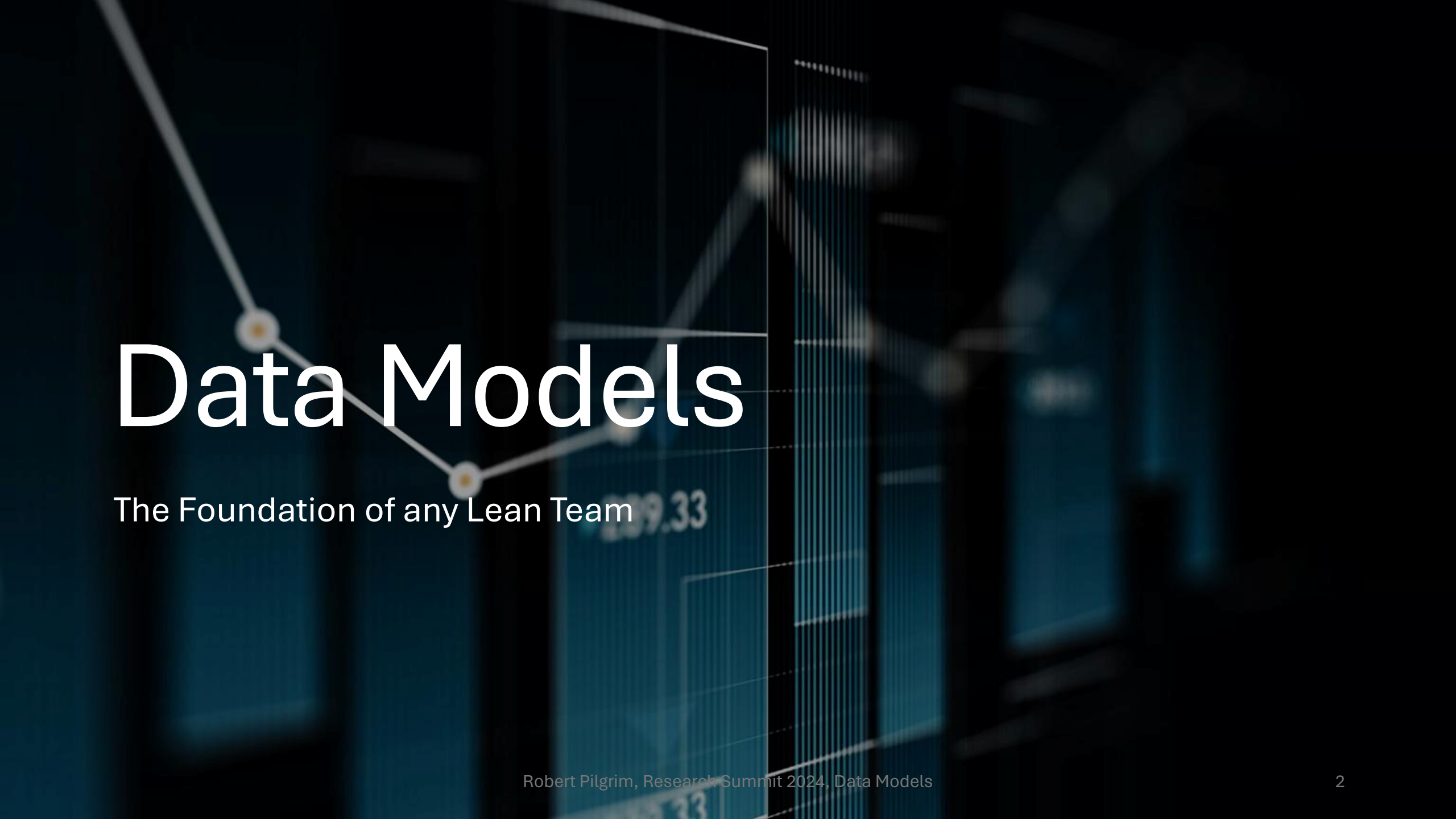
2024
Research Analytics Summit
Building Capacity, Empowering Skill Transfer, and Energizing a Growing Community

Data Models

Robert Pilgrim
rpilgrim@uark.edu

Check out my blog

Connect with me



Data Models

The Foundation of any Lean Team

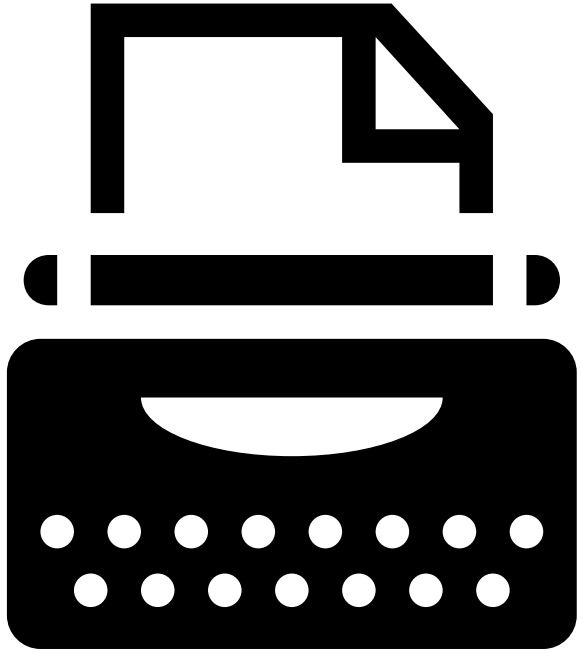
Who am I

Robert Pilgrim, rpilgrim@uark.edu

Associate Director of Data Strategy & Insights

Division of Research & Innovation, University of Arkansas

Overview

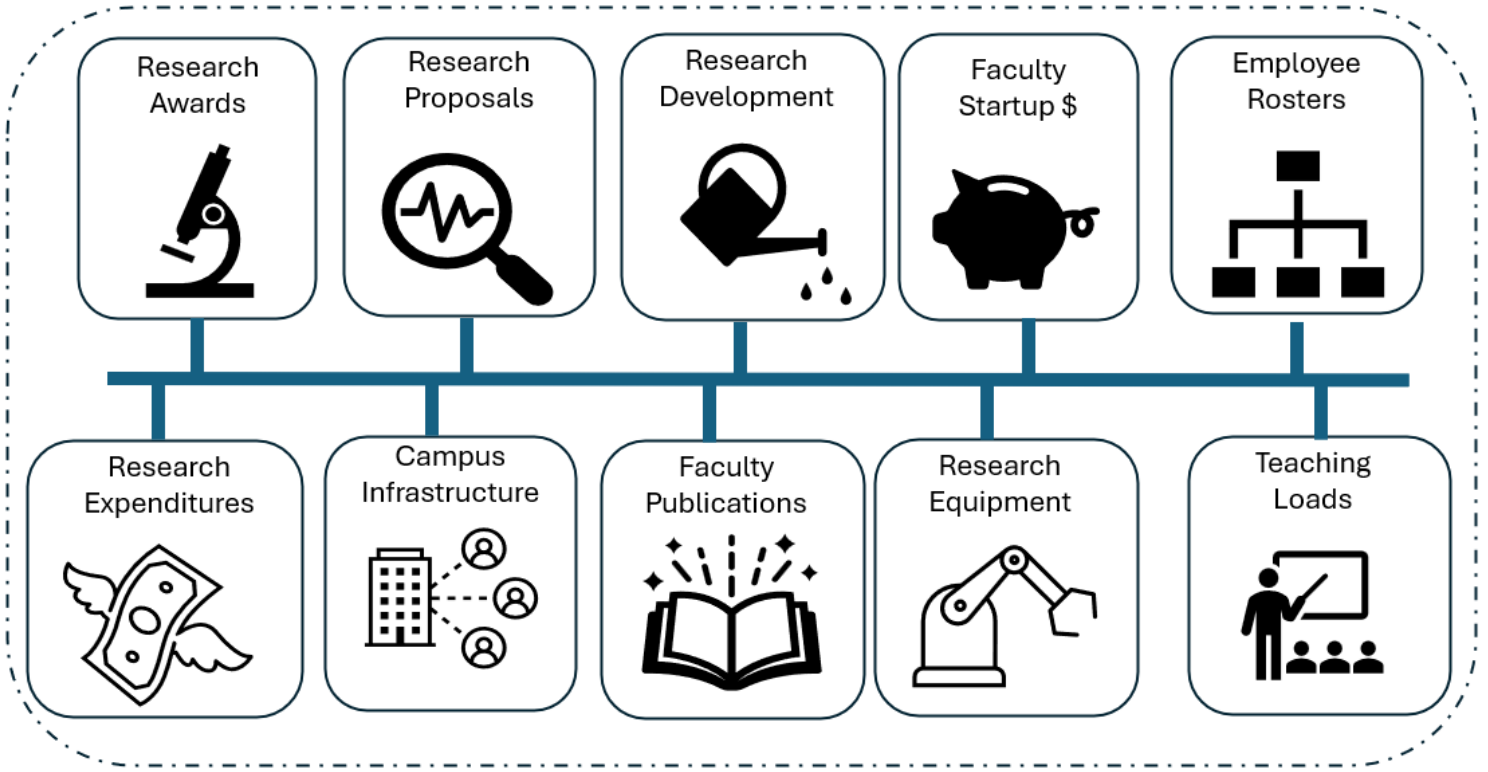


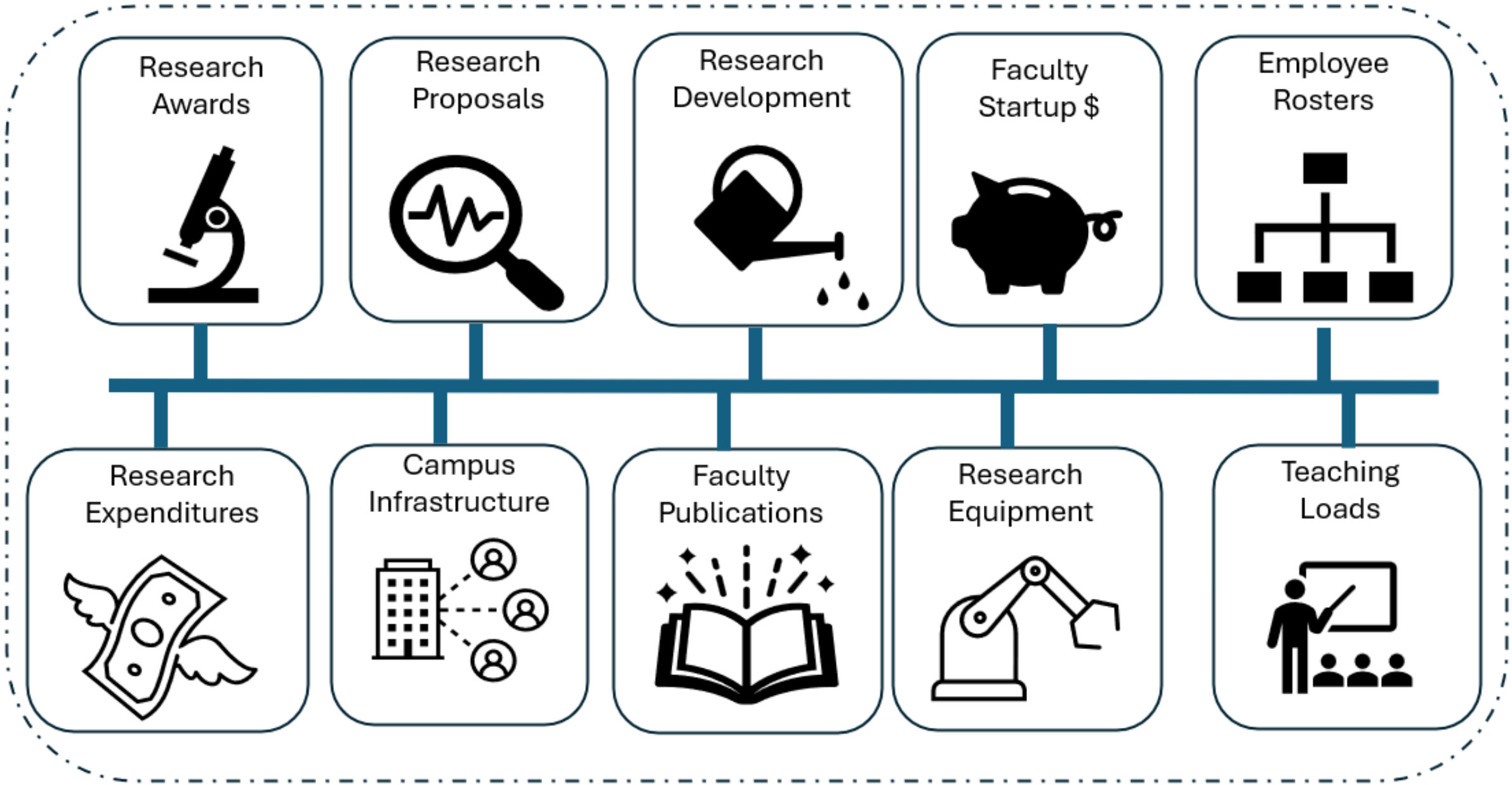
- What is a Data Model (DM)
- Benefits of Using a DM
- Focus on Sponsored Research related data
  - Campus Strategic Model
- Cross-Data calculations
  - DAX (Data Analysis Expressions) Language
- Pivot Tables
- Building with your Data Model

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




Research Awards
Research Proposals
Research Development
Faculty Startup $
Employee Rosters
Research Expenditures
Campus Infrastructure
Faculty Publications
Research Equipment
Teaching Loads

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

1 Raw Data

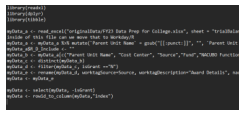
2 Code

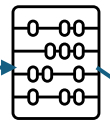
3 Calculations

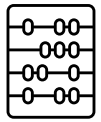
4 Merge

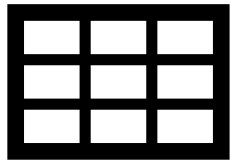
Awards Data

Proposal Data









Merging
Vlookups
SQL, R, SAS

5

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

Next : More Realistic Example

More Realistic Example

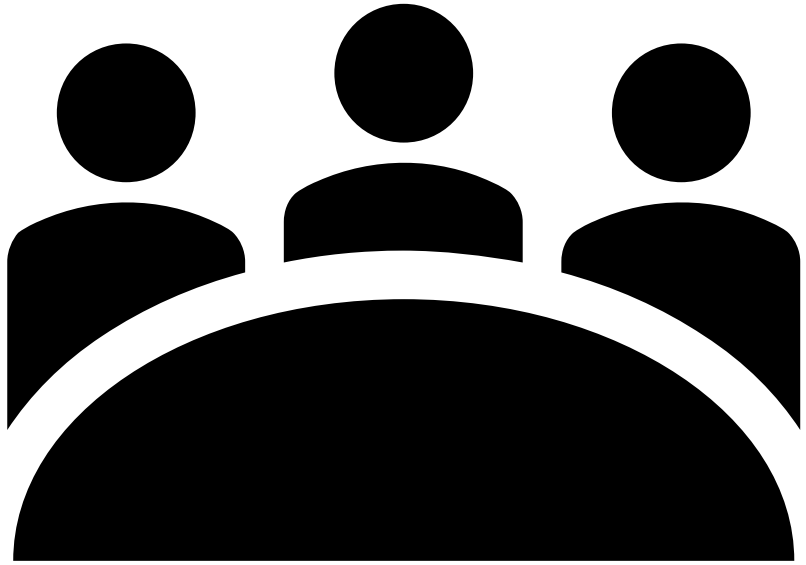
Initial Request | Campus

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



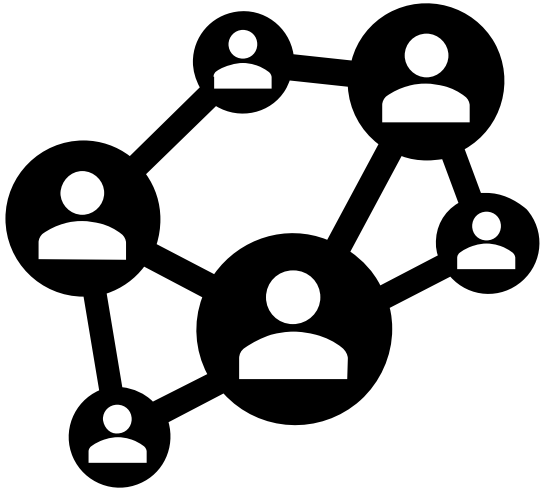
More Realistic Example

Initial Request | Campus

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

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

By College



Remove People

Exclude Sponsors

My calculations don't match yours?

More Realistic Example

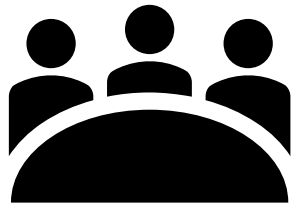
Initial Request | Campus

Award $

Proposal Success Rates,

Proposal Counts,

Total unique PI's



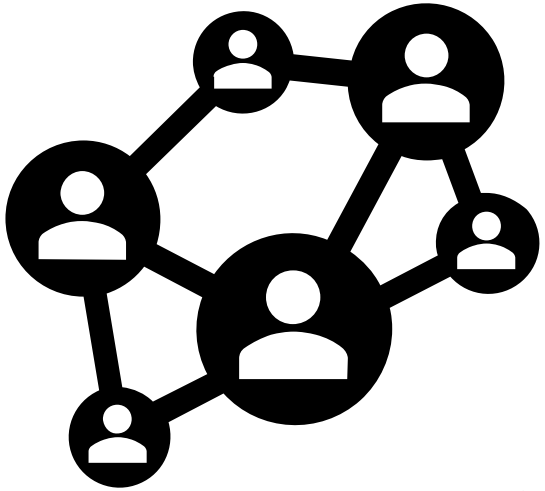
Faculty Level

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

By College

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

Remove People



How do Tenure Track vs Tenured faculty metrics compare?

Exclude Sponsors

Where is the most recent version of the data?

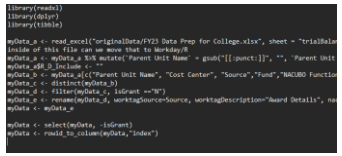
Can you put this into a line chart and do some trend analysis

My calculations don't match yours?

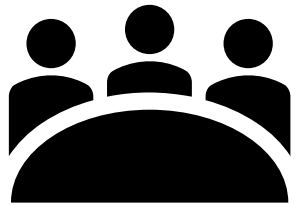
Multiple Code

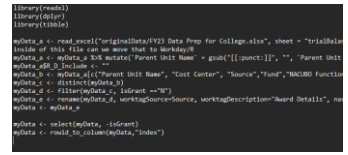
Initial Request | Campus

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

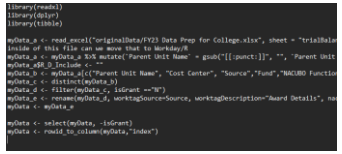


Faculty Level





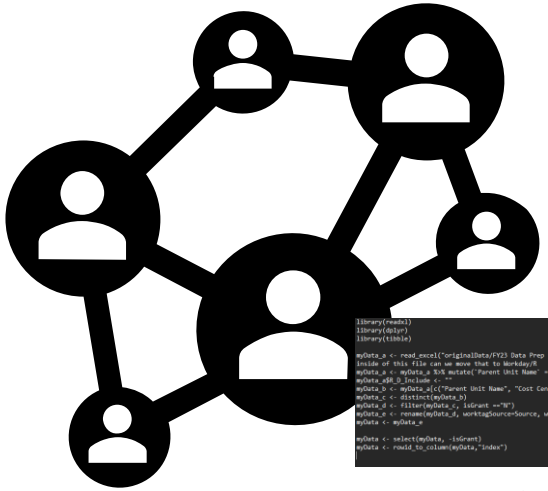
Can you send the merged data to X so they can build some reports



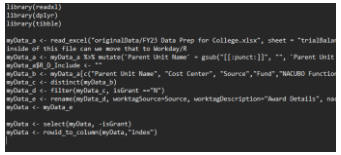
By College

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

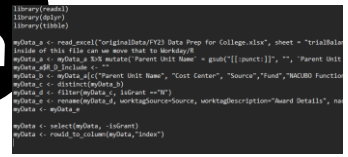
Remove People



How do Tenure Track vs Tenured faculty metrics compare?



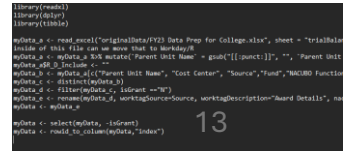
Where is the most recent version of the data?



Exclude Sponsors

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

My calculations don't match yours?



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 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%**

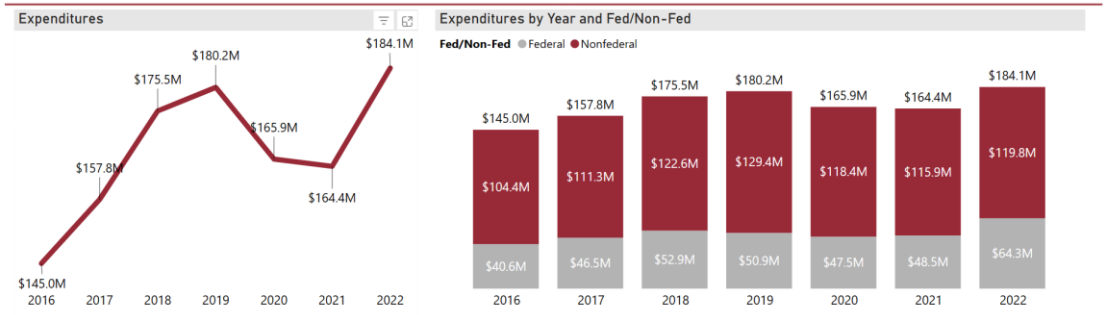
Separate Reporting from Data Model

Data Model

- Design
- Create

Reporting

- Create
- Distribute

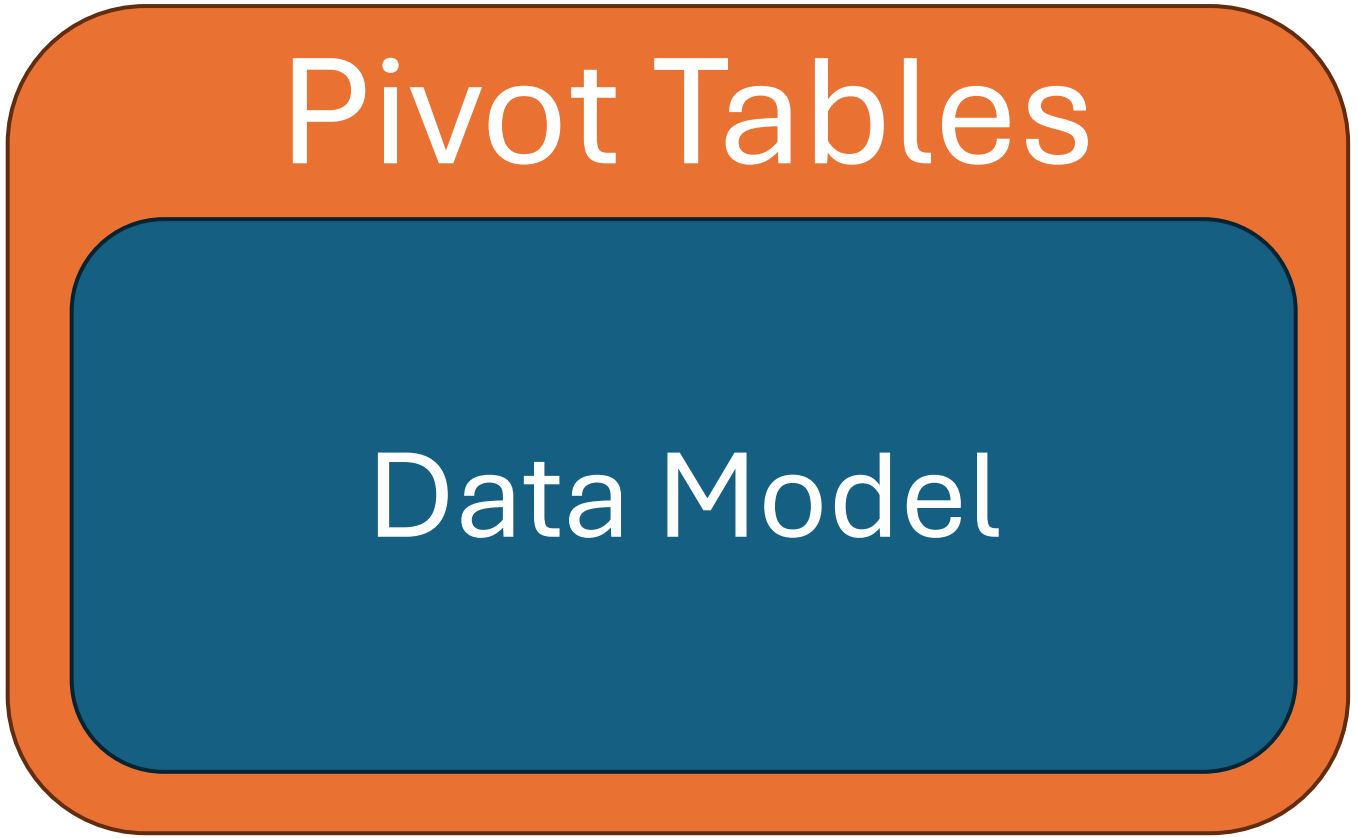


Fed/Non-Fed	2016	2017	2018	2019	2020	2021	2022	Total
Federal	$40,642,000	$46,490,000	$52,937,000	$50,875,000	$47,454,000	$48,534,000	$64,252,000	**$351,184,000**
Nonfederal	$104,378,000	$111,301,000	$122,561,000	$129,350,000	$118,433,000	$115,909,000	$119,828,000	**$821,760,000**
Total	**$145,020,000**	**$157,791,000**	**$175,498,000**	**$180,225,000**	**$165,887,000**	**$164,443,000**	**$184,080,000**	**$1,172,944,000**

Next: Pivot Table Image

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

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**

Fiscal Year

Investigator

Sponsor

College

Research Proposals

Proposal **Number**

Fiscal Year

Proposal **Title**

Proposal **PI**

Proposal **Investigator**

Proposal **Investigator Role**

Proposal **Date**

Proposal **Investigator College**

Proposal **Sponsor**

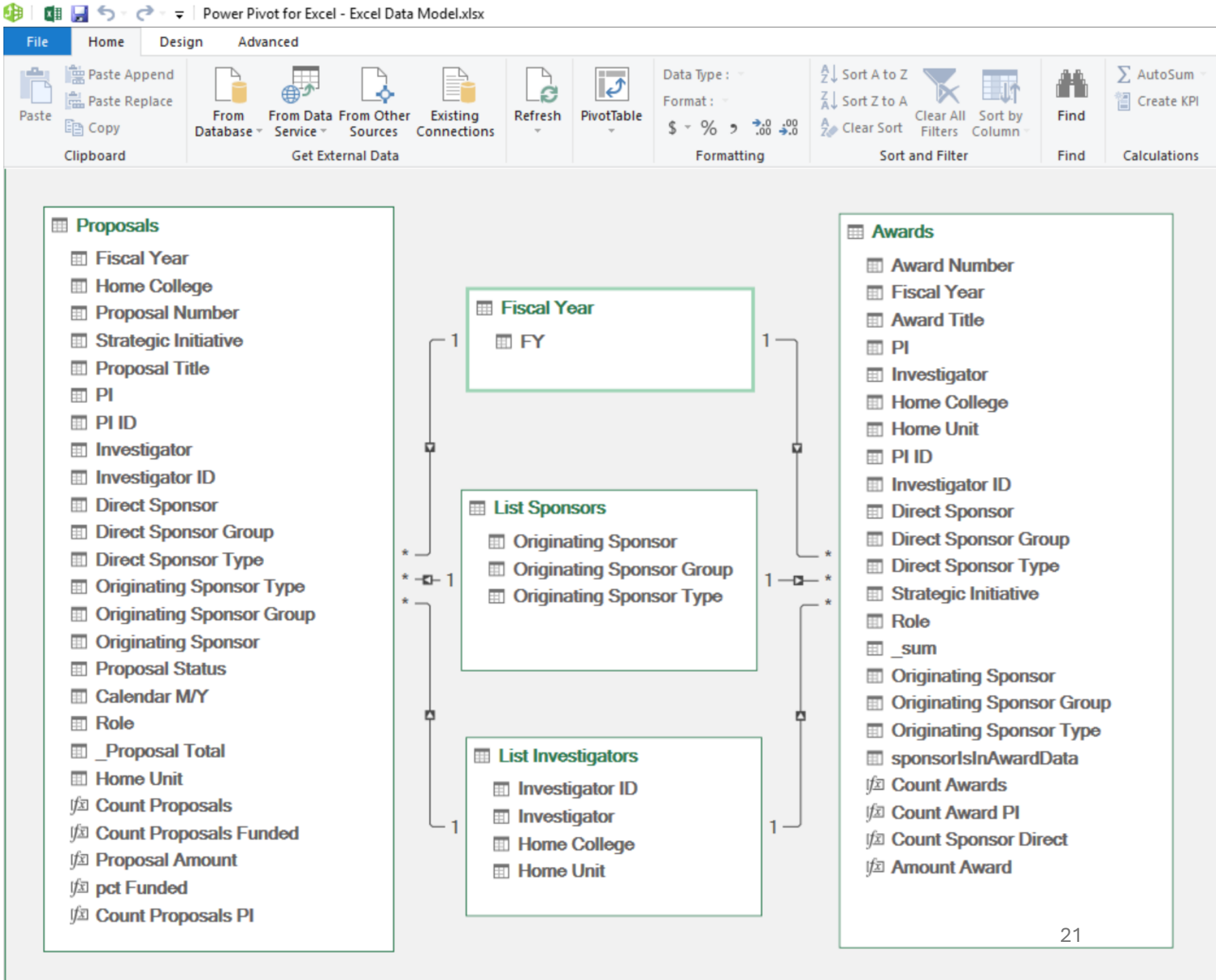
Proposal **Sponsor Type**

Proposal **Sponsor Type Group**

Proposal **Amount**

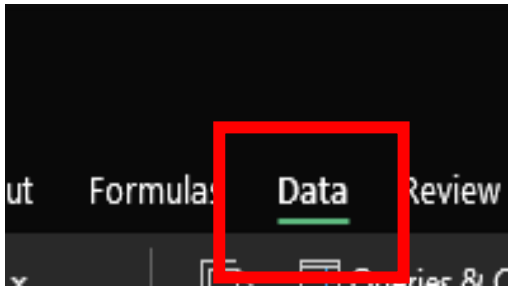
Proposal **Status**

Excel Data Model

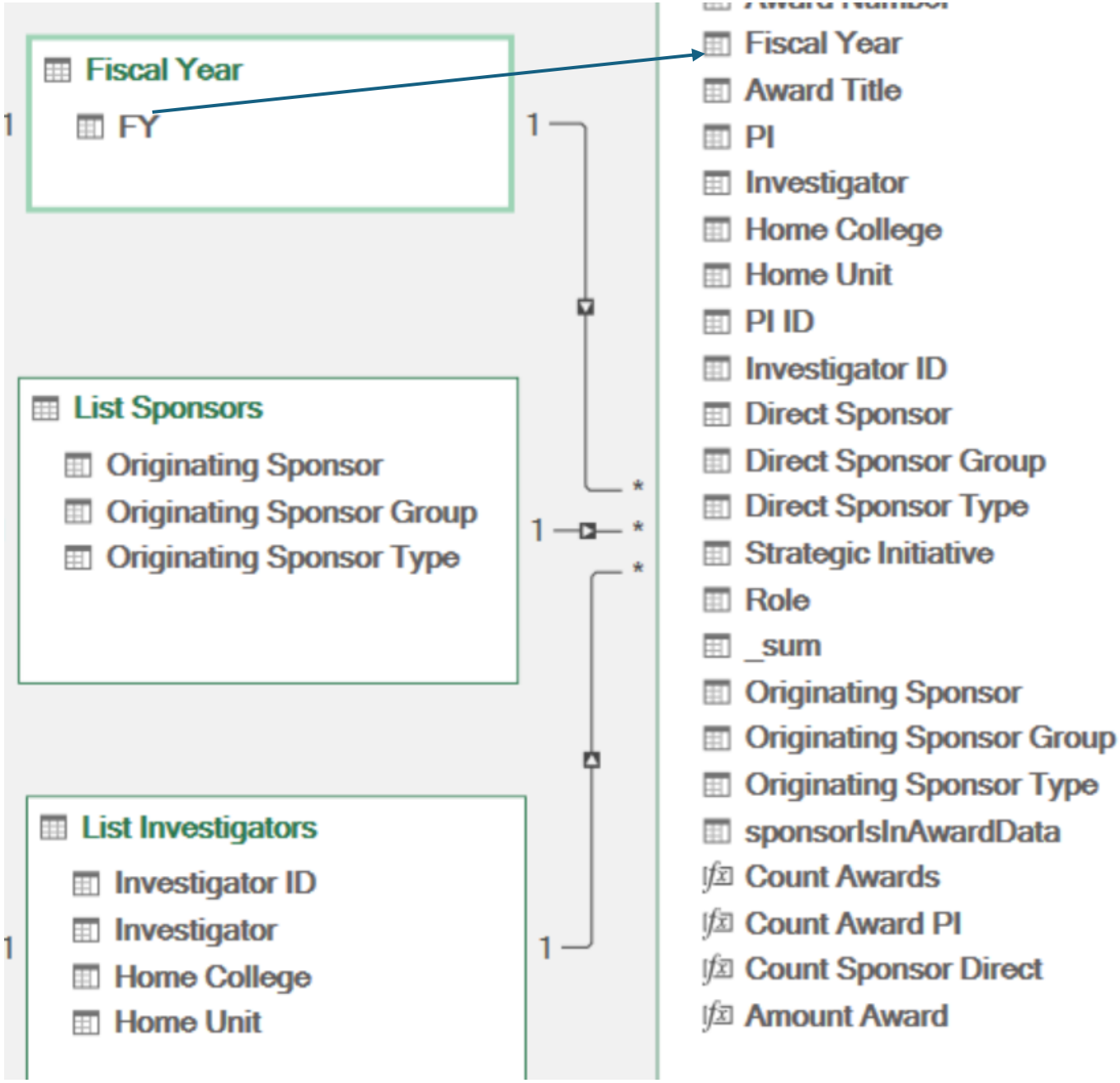


Next : Excel DM

Excel Data Model



Drag Connecting Variables onto Matching variables in



Next: Data Model Calculations

Data Model Calculations

Custom Built Functions for Cross-Data Analytics

Hello World | DAX Language

Data Analysis Expressions (DAX)

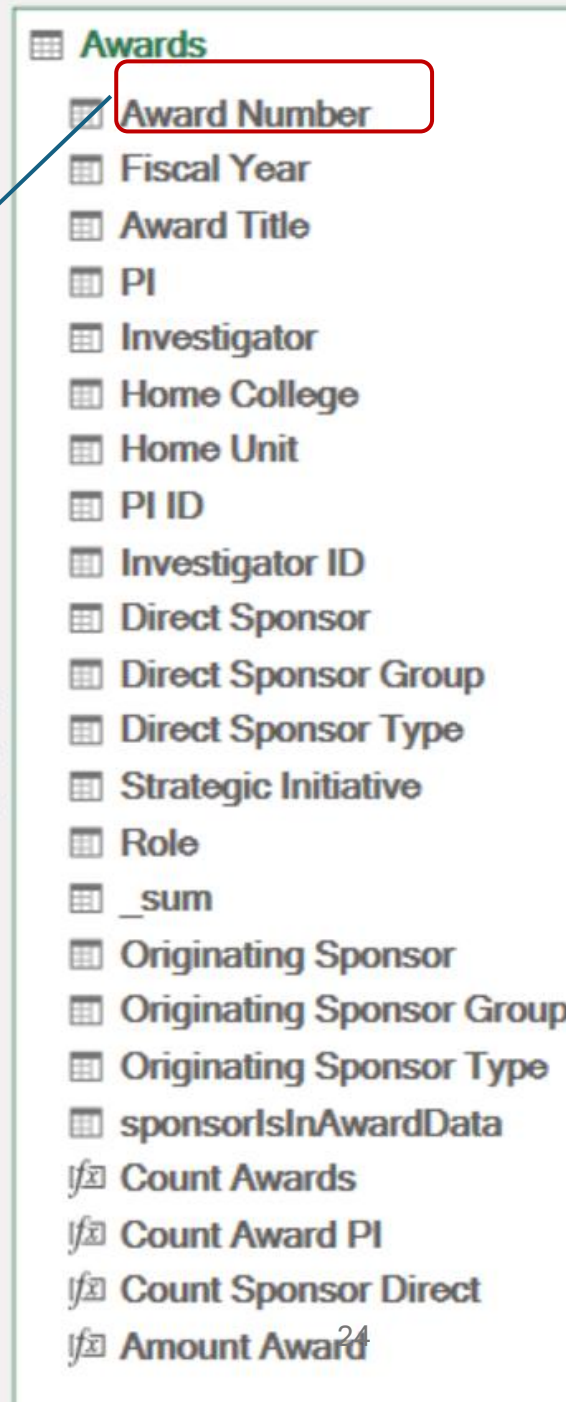
Calculation Name

Table

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

Function

Variable



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])
```

Next : Cross-Data Calculations

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]

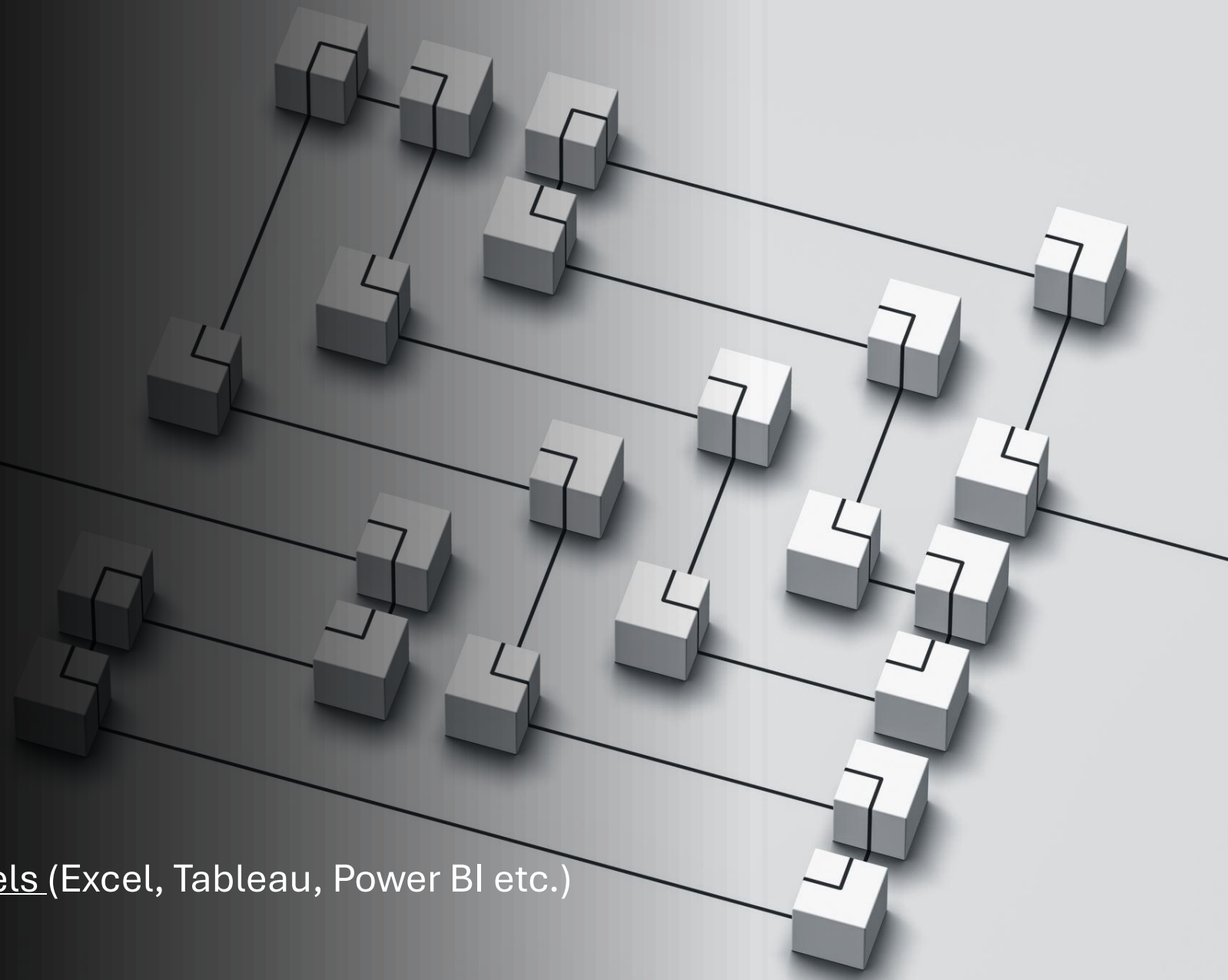
4. Avg Proposal Submission Rate per PI
[Count Proposals]/ [Count PI]

Bringing this all Together

Data Model + Pivot Table

Next : Pivot Table Concept

The Concept of a Pivot Table

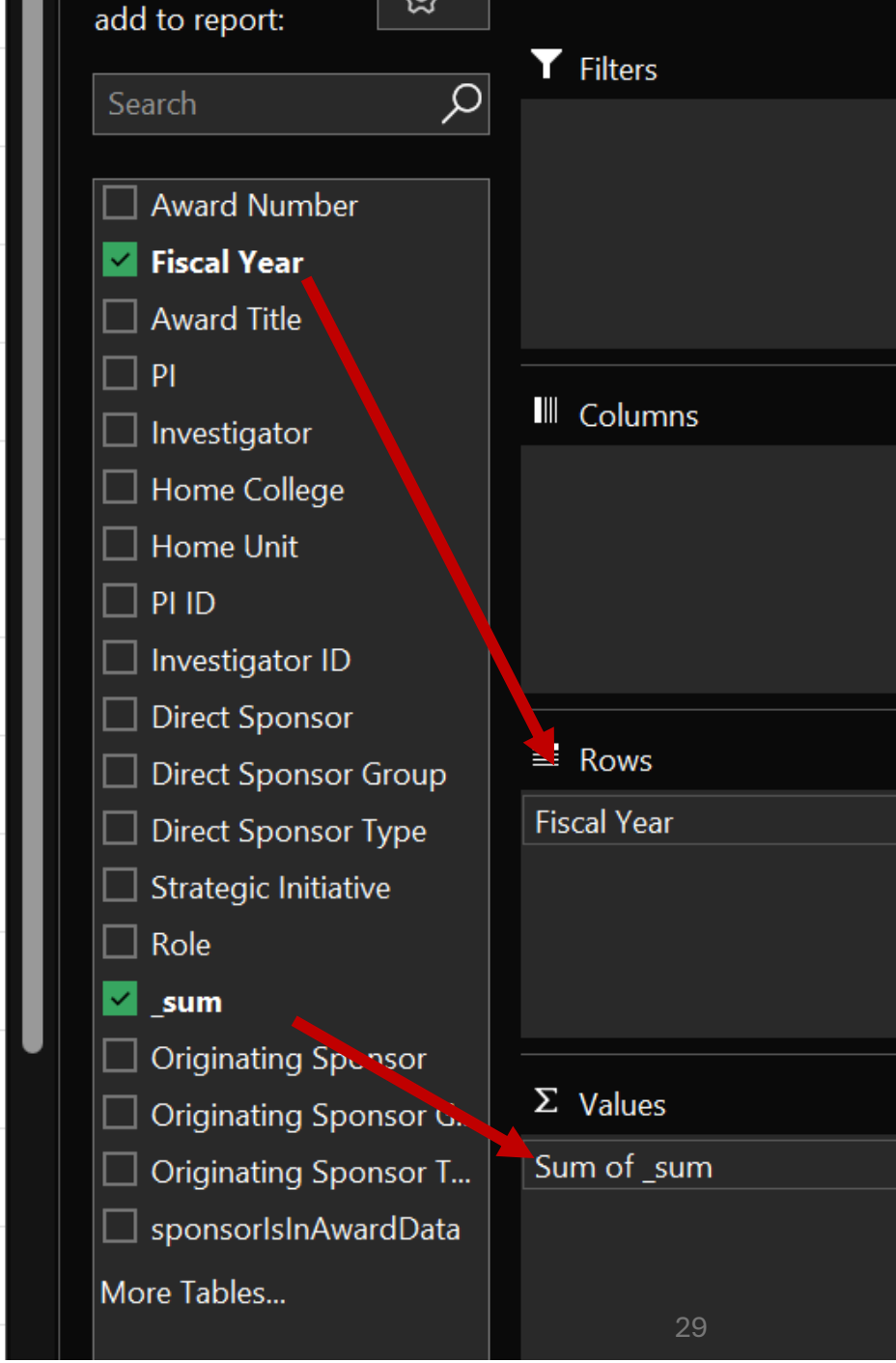


Foundation Element of All Data Models (Excel, Tableau, Power BI etc.)

Pivot Tables 101

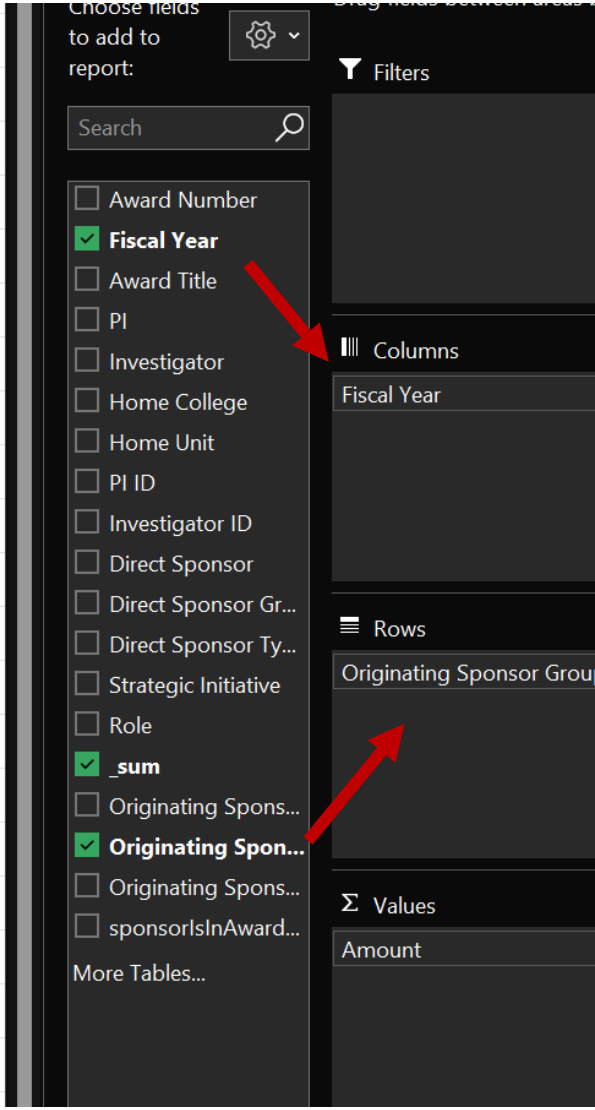
Row Labels	Sum of _sum
2024	$250,351,073
2023	$447,608,567
2022	$438,148,800
2021	$317,595,905
2020	$255,016,878
2019	$276,841,837
2018	$205,866,572
Grand Total	**$2,191,429,632**

Aggregation Calculations e.g. SUM



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**



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"

Next : Drum Roll

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]

Cross Data Calculations

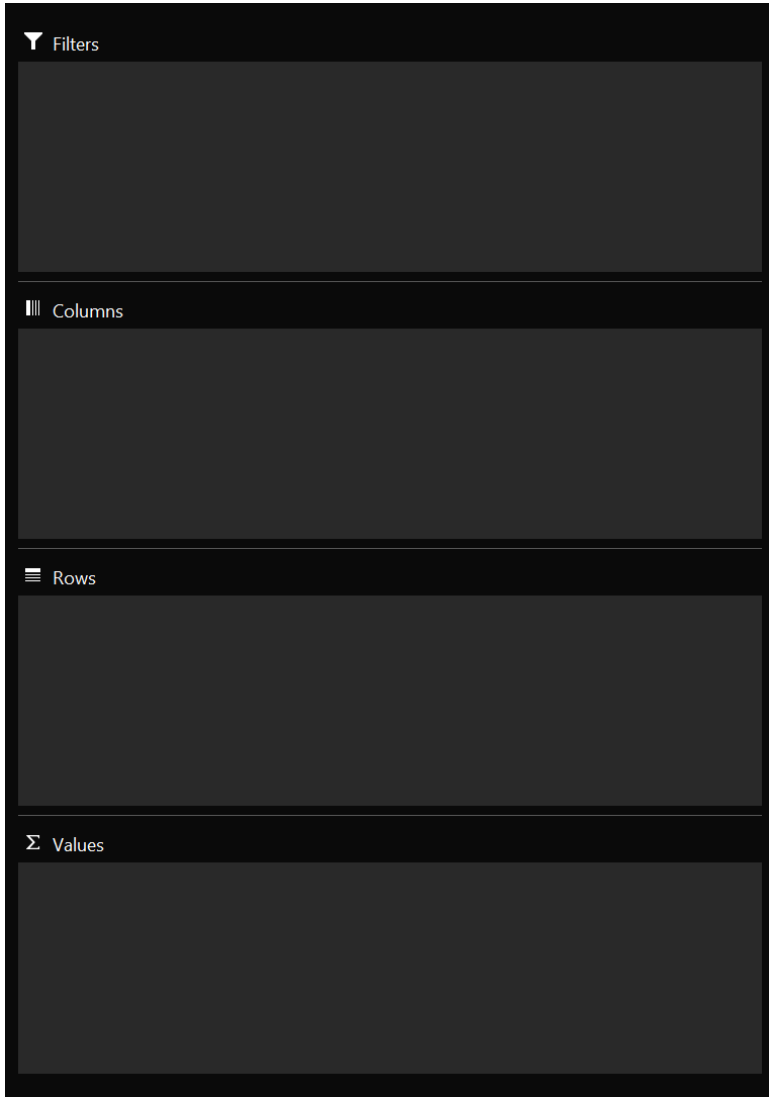
Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

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



Unit

Emp ID

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

[Count Proposals Funded]

[Count PI]

[Research Exp.]

[Research Area]

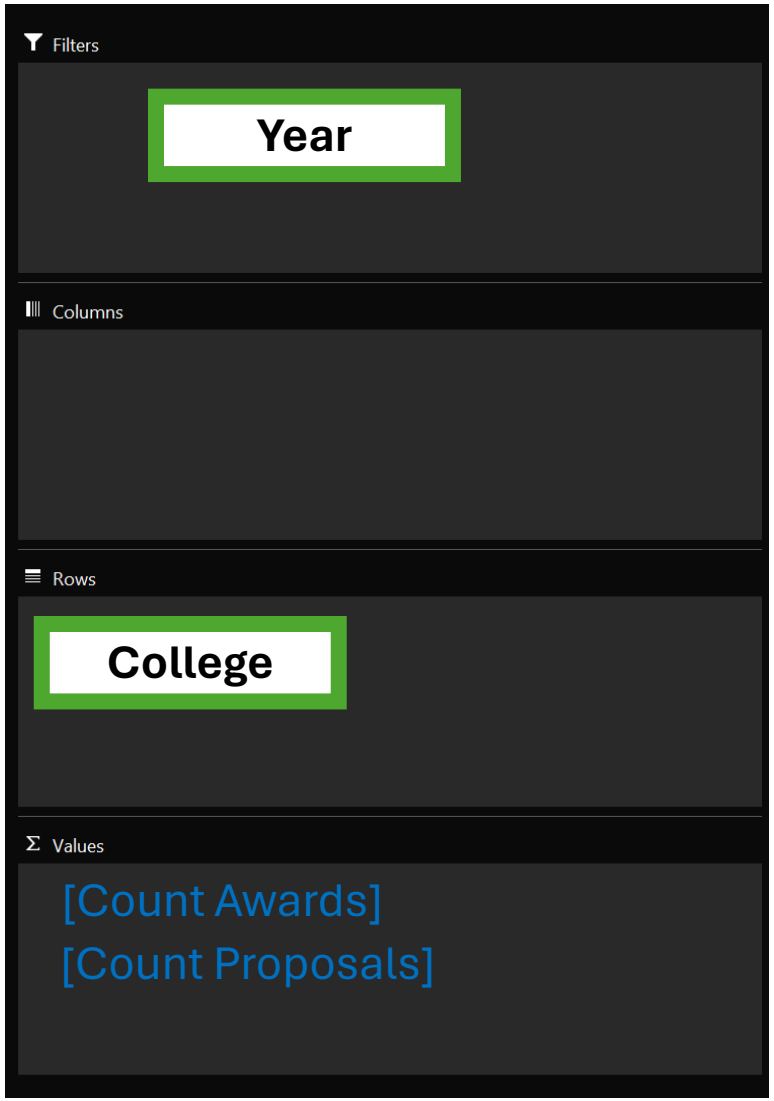
Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI



College | **Unit** | **Year** | **Emp ID**

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Cross Data Calculations

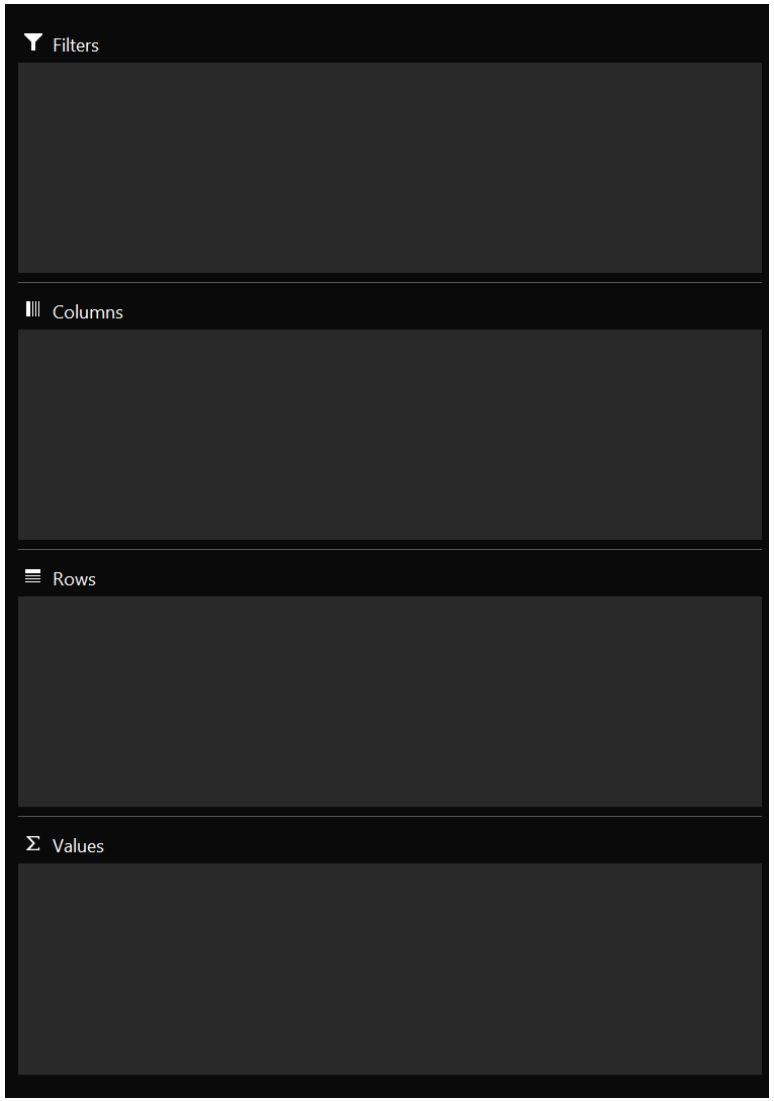
Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

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



Unit

Emp ID

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

[Count Proposals Funded]

[Count PI]

[Research Exp.]

[Research Area]

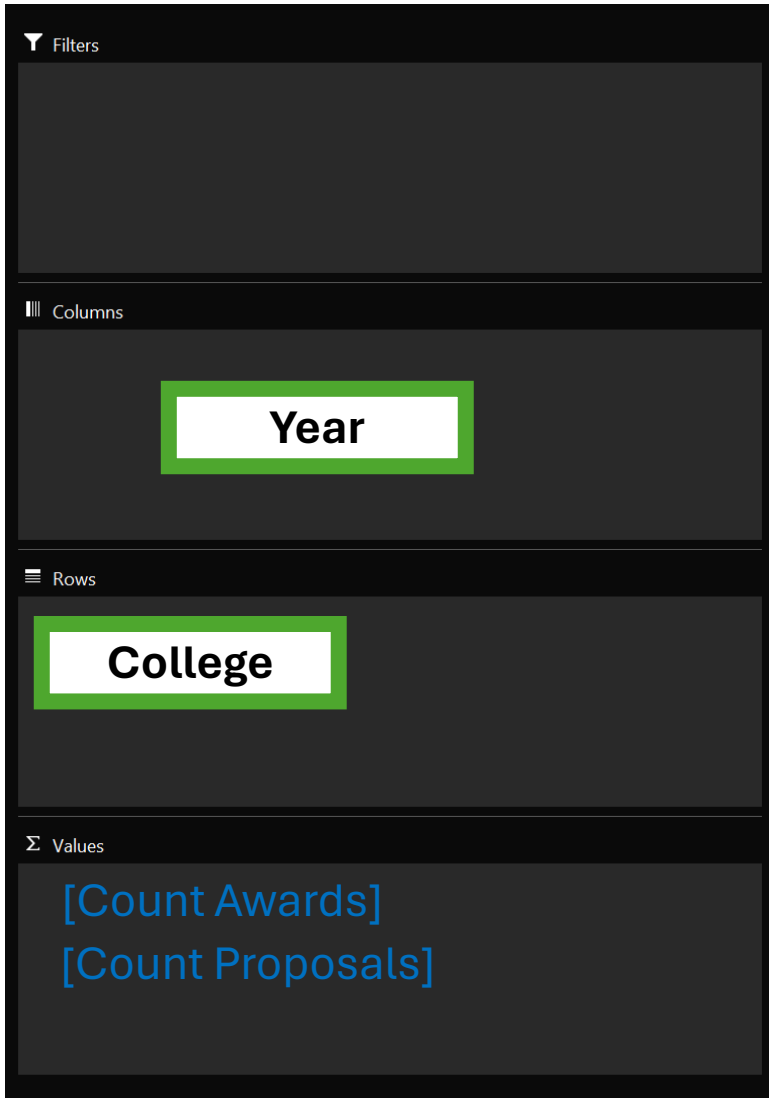
Cross Data Calculations

Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI



College | **Unit** | **Year** | **Emp ID**

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Cross Data Calculations

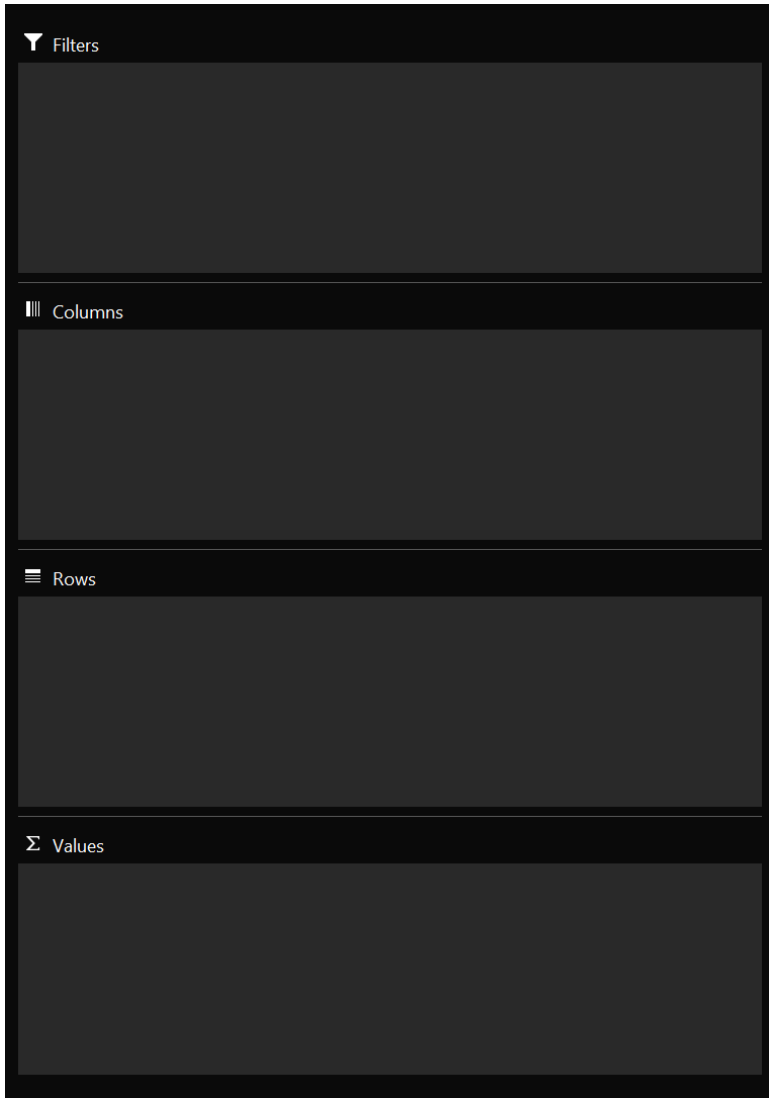
Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

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



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]

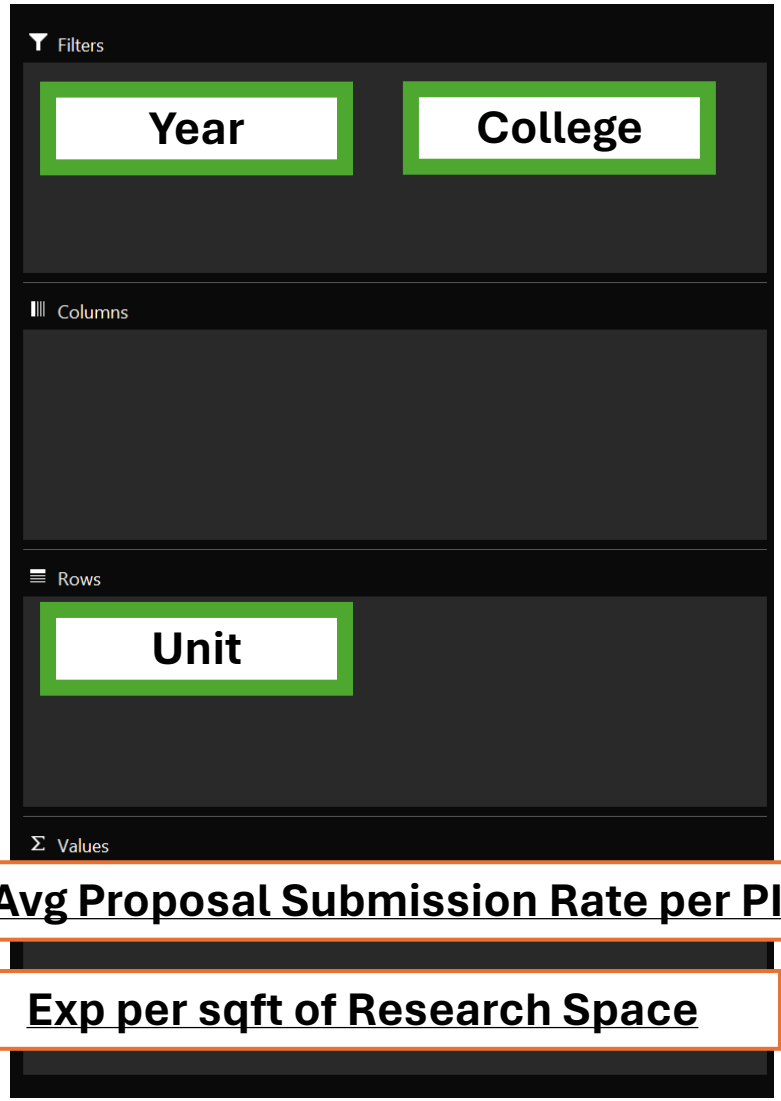
[Research Exp.]

[Research Area]

Cross Data Calculations

Proposal Success Rate

Avg Exp per PI



College | **Year** | **Unit** | **Emp ID**

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

Cross Data Calculations

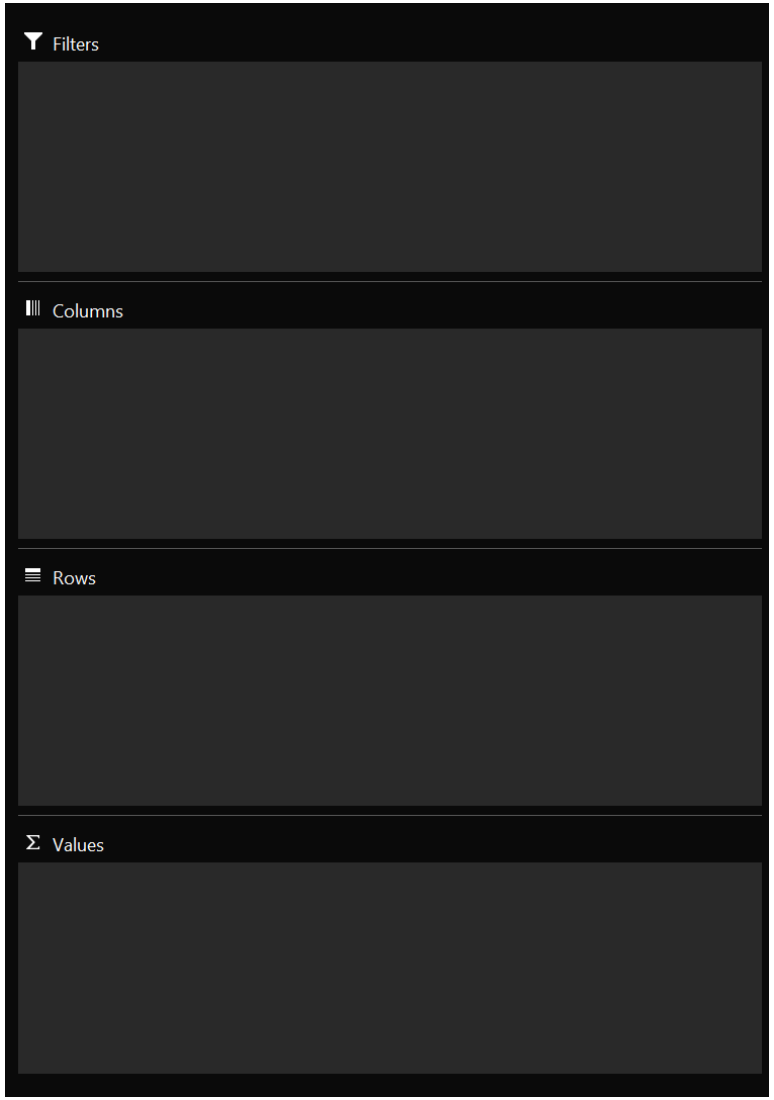
Proposal Success Rate

Exp per sqft of Research Space

Avg Exp per PI

Avg Proposal Submission Rate per PI

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



Unit

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

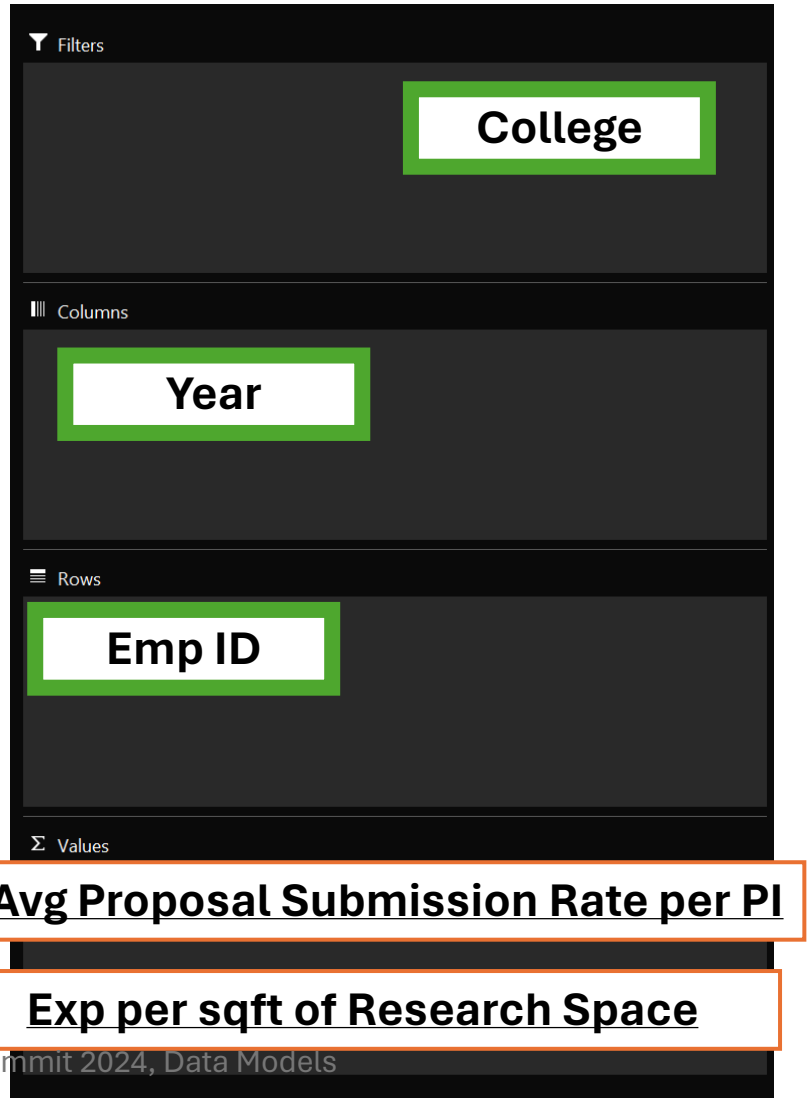
[Research Area]

Cross Data Calculations

Proposal Success Rate

Avg Exp per PI

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



All Metrics by Unit and Year

Row

Unit

Aggregation

[Count Proposals]

[Count Proposals Funded]

[Count PI]

[Count Awards]

[Research Exp.]

[Research Area]

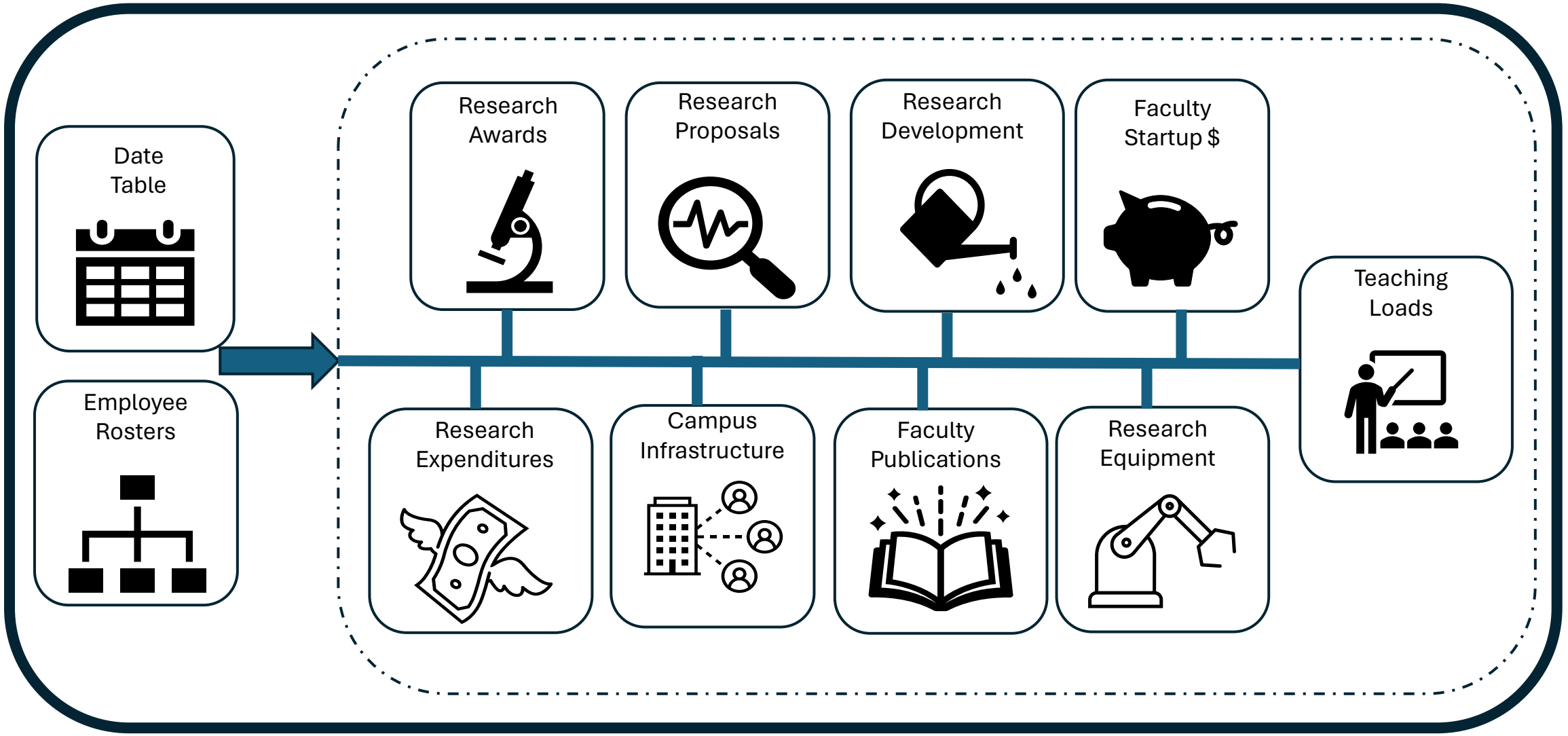
Proposal Success Rate

Avg Exp per PI

Avg Proposal Submission Rate per PI

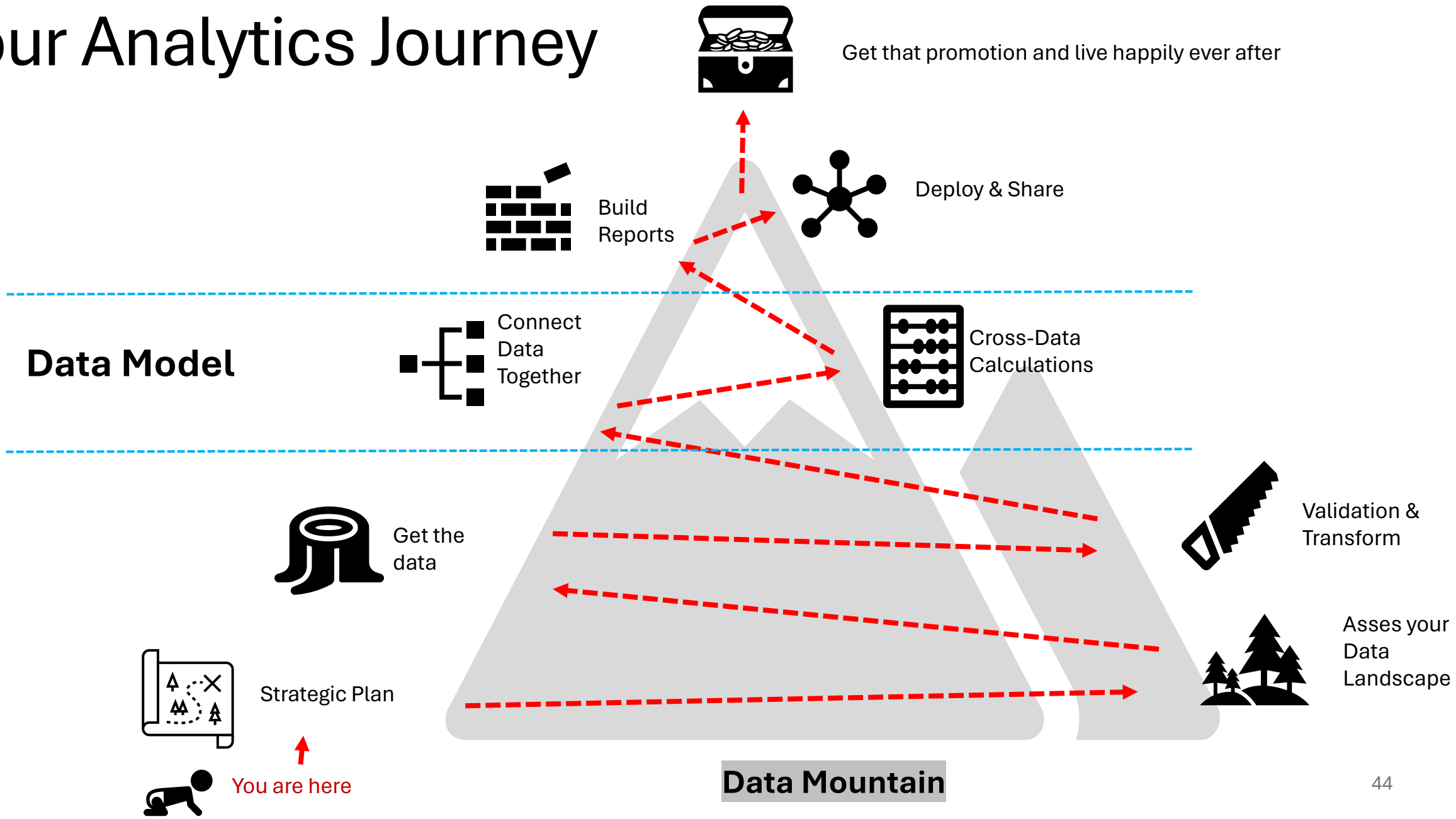
Exp per sqft of Research Space

Typical University Data Model (Research Focused)



>>Last Slide

Your Analytics Journey



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