



# DATA WRANGLING AND VISUALIZATION IN ACCESS SERVICES



Ellie Kohler, MA, MLIS  
Access and Learning Services Librarian  
Rockhurst University Library  
November 17, 2016

2016 Access Services Conference  
Atlanta, Georgia



[HTTP://TINY.CC/DATAWRANGLE](http://tiny.cc/datawrangle)

**Links to some of the tools I'm talking about  
today.  
Your suggestions!**

Disclaimer:

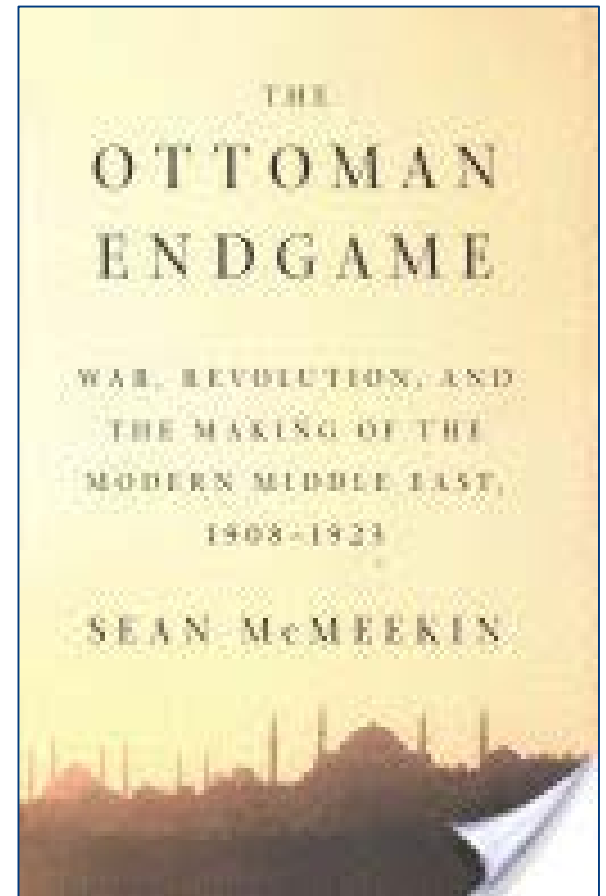
This presentation will not make you an expert.



# You know data

---

- DS 63.04 .M38 2015
- The Ottoman Endgame
- About the Ottoman Empire
- It's yellow





# Where do you get your data?

---

- Google sheets: internal reports
- Web Management Reports
- ILS System: III (Sierra)
- MOBIUS Statistics
- OCLC Web Reports
- Wufoo
- Gimlet
- The list goes on...

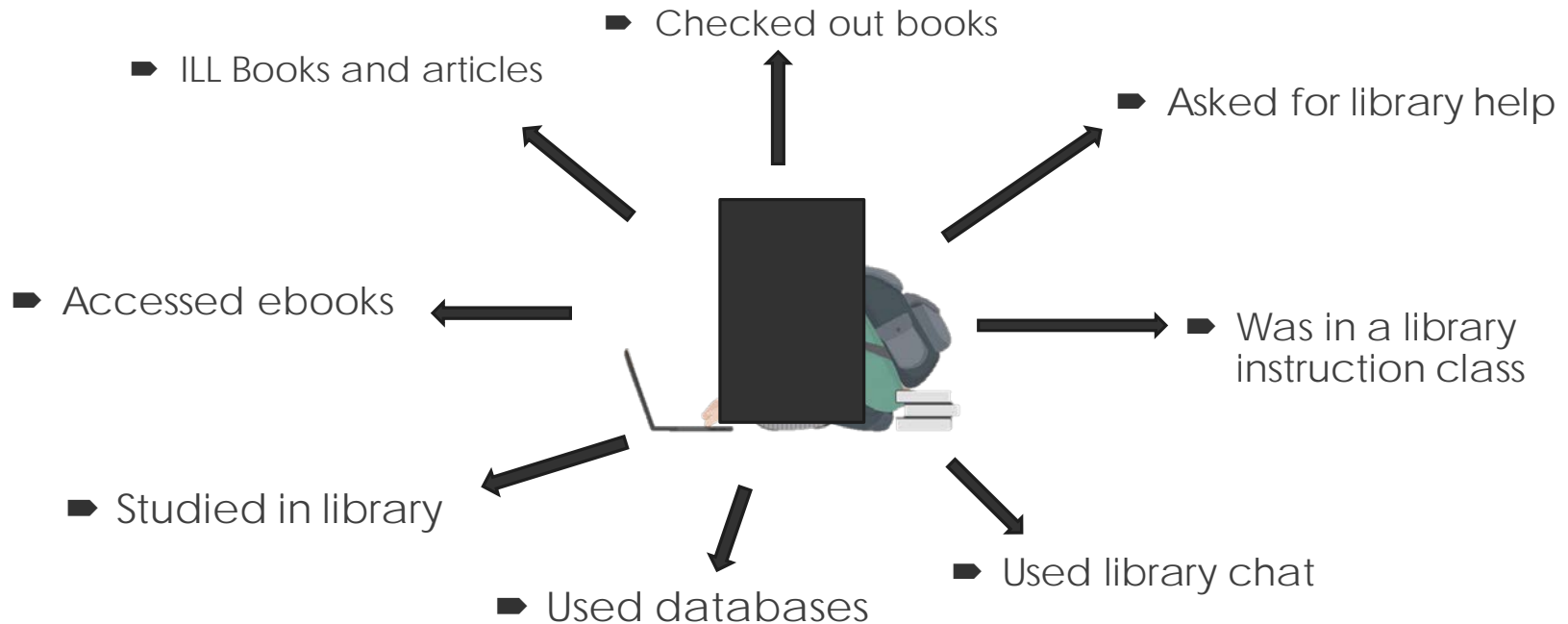




**All identifiable information  
is blocked or stripped from  
these transactions.**

Measures of Student Success

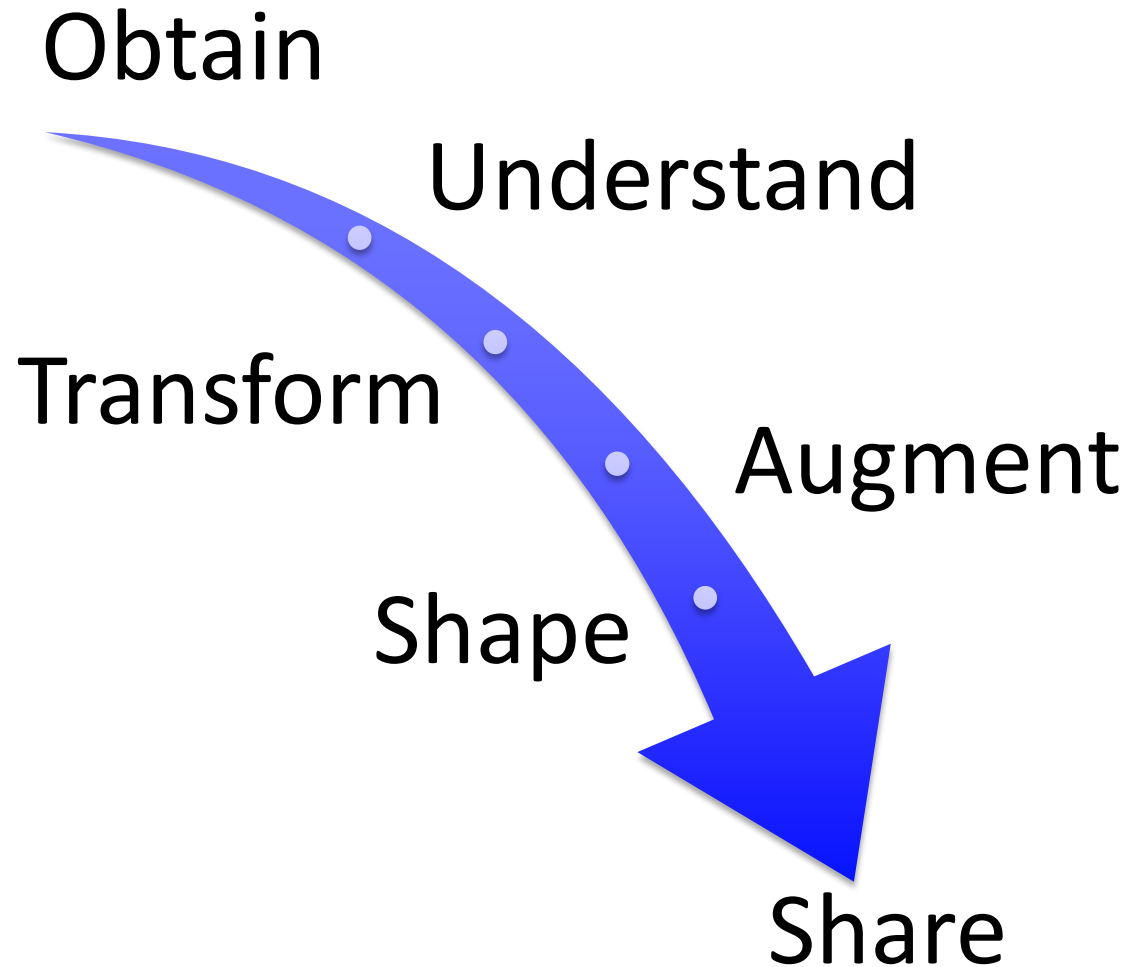
- ▀ Grades
- ▀ Graduation





# What is Data Wrangling?

---





# Hell is other people's data

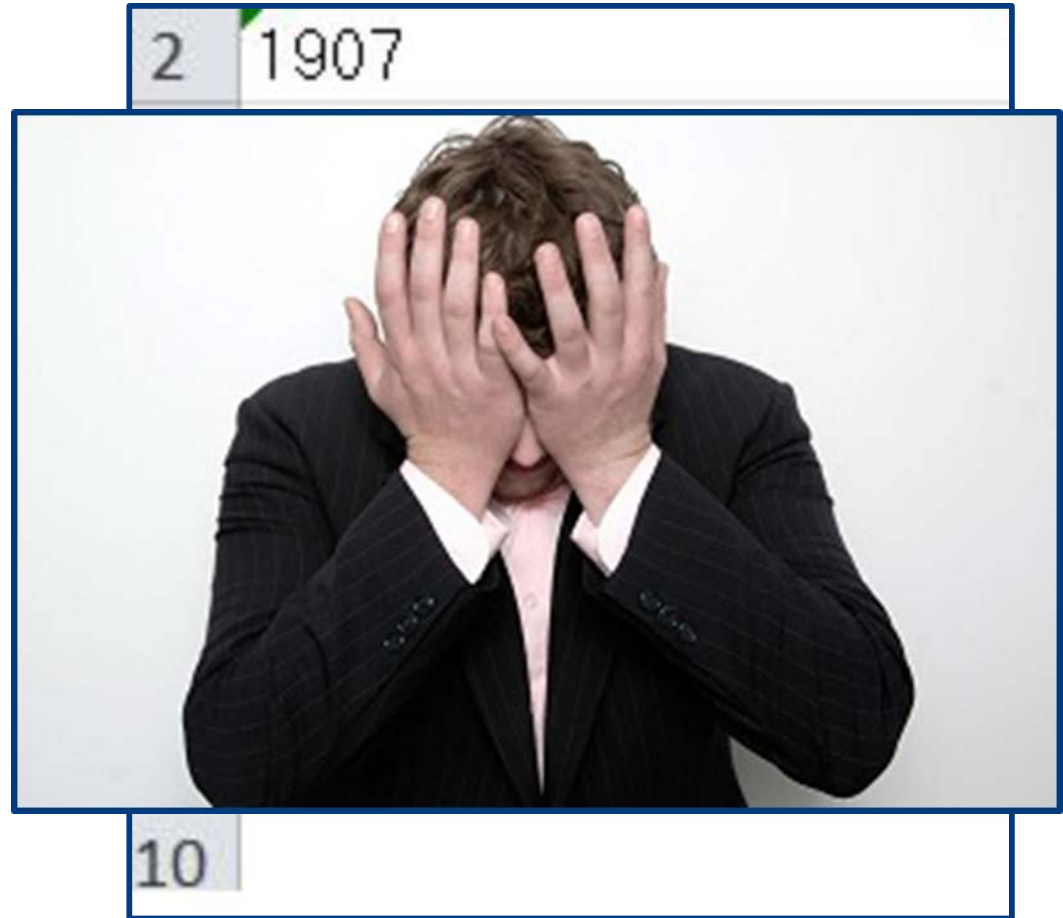
---





# 80 / 20 Rule

- Languages
- Dates
- “Revisions”
- Mistakes- Typos
- Blanks
- Aliasing
- Wrong format
- Too specific
- Too broad





# Free Tools That Can Help

---



- Free
- Download



- Free version
- Online
- Bigger datasets



- Free version
- Download
- Version 1.3

# Open Refine: Open Text Fields

**Greenlease Library** |  **ROCKHURST UNIVERSITY**

## Interlibrary Loan: Journal Article Request

Fill out the form below to request journal articles not available at the library.  
Questions? Go to the [Interlibrary Loan guide](#) or contact a librarian at 816-501-4116.

**Article Title \***

**Author(s) \***

**Complete Journal Title \***

**Publication Year or Full Date \***

**Volume \*** **Issue \*** **Page numbers \***

*Example is bolded:* 78(8):e54-58      *Example is bolded:* 78(8):e54-58      *Please include the beginning and ending page numbers:*

*Example is bolded:* 78(8):e54-58


**Database or website where citation was found?**

**Complete Journal Title** change

2604 choices **Sort by: name count** Cluster

- Annals of Clinical Psychiatry 1
- Annual review of medicine 1
- Arthritis and Rheumatology 1
- Augmentative and alternative communication ( 1
- Australian Occupational Therapy Journal 1
- Biology of Blood and Marrow Transplantation 1
- Brain injury 1
- British Journal of Psychiatry 1
- Burns 1
- Canadian journal of occupational

**Complete Journal Title** change reset



1.00 — 44.00

**Article 1**

Implementa  
f an electr  
resource  
ssessment  
system in a  
ademic  
brary.

constraint-  
duced  
ovement  
therapy for  
children W  
lemiplegia  
fter Trauma  
rain Injury  
quantitati  
study

ffect of  
eatment  
nvironmen  
modified  
onstraint-  
duced  
ovement  
therapy res  
children  
pastic  
emiplegic  
problema



## Cluster & Edit column "Complete Journal Title"

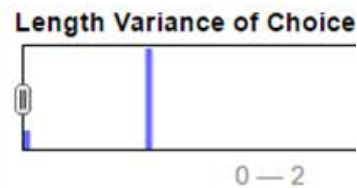
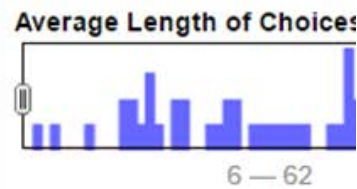
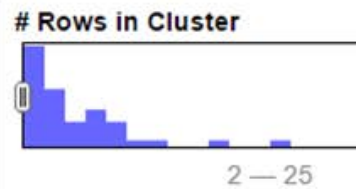
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find](#)

Method key collision

Keying Function fingerprint

44

2	2	<ul style="list-style-type: none"><li>International Review for the Sociology of Sport (1 rows)</li><li>International Review for the Sociology of Sport (1 rows)</li></ul>	<input checked="" type="checkbox"/>	International Review for the Soci
2	6	<ul style="list-style-type: none"><li>Neuropsychologia (4 rows)</li><li>Neuropsychologia (2 rows)</li></ul>	<input checked="" type="checkbox"/>	Neuropsychologia
2	2	<ul style="list-style-type: none"><li>Journal of Research in Childhood Education (1 rows)</li><li>Journal of Research in Childhood Education (1 rows)</li></ul>	<input checked="" type="checkbox"/>	Journal of Research in Childhood
2	4	<ul style="list-style-type: none"><li>Equity and Excellence in Education (3 rows)</li><li>Equity and Excellence in Education (1 rows)</li></ul>	<input checked="" type="checkbox"/>	Equity and Excellence in Educati
2	2	<ul style="list-style-type: none"><li>BioEssays (1 rows)</li><li>Bioessays (1 rows)</li></ul>	<input checked="" type="checkbox"/>	BioEssays
2	8	<ul style="list-style-type: none"><li>Psychiatry Research (7 rows)</li><li>Psychiatry Research (1 rows)</li></ul>	<input checked="" type="checkbox"/>	Psychiatry Research
2	3	<ul style="list-style-type: none"><li>Brain Research Bulletin (2 rows)</li><li>Brain research bulletin (1 rows)</li></ul>	<input checked="" type="checkbox"/>	Brain Research Bulletin



# When clustering fails you....

Refresh Reset All Remove All

☒ Year Only change

685 choices Sort by

Method nearest neighbor Distance Function levenshtein Radius 1.0 Block Chars 6 185 clusters found

02/01/1989) 1  
(04/15/2007 1  
(05/31/2010). 1  
01/01/1988 2  
01/01/1996 2  
01/01/1999 2  
01/1993 2  
01/2009 2  
02/01/14 1  
02/01/1996 1  
02/01/2006 1  
02/01/2007 1

Nov 2009 (4 rows)  
Nov 2008 (3 rows)  
Nov 2007 (3 rows)  
Nov 2001 (3 rows)  
Nov 2000 (2 rows)

8 19 January 2011

January 2011 (6 rows)  
January 2004 (3 rows)  
January 2005 (3 rows)  
January 2009 (2 rows)  
January 2007 (2 rows)  
January 2001 (1 rows)  
January 2003 (1 rows)  
January 2006 (1 rows)

8 14 January 2004

January 2004 (3 rows)  
January 2005 (3 rows)  
January 2009 (2 rows)  
January 2007 (2 rows)  
January 2000 (1 rows)  
January 2001 (1 rows)  
January 2003 (1 rows)  
January 2006 (1 rows)

7 18 Aug 2010

Aug 2010 (6 rows)  
Aug 2005 (3 rows)  
Aug 2000 (2 rows)  
Aug 2003 (2 rows)  
Aug 2002 (2 rows)

Select All Deselect All Merge Selected & Re-Cluster Merge Selected & Close Close

# Choices in Cluster

2 — 9

# Rows in Cluster

2 — 38

Average Length of Choices

8 — 16

Length Variance of Choices

0 — 0.5

# R: Free Tool for Coding



## Packages:

dplyr

tidyr

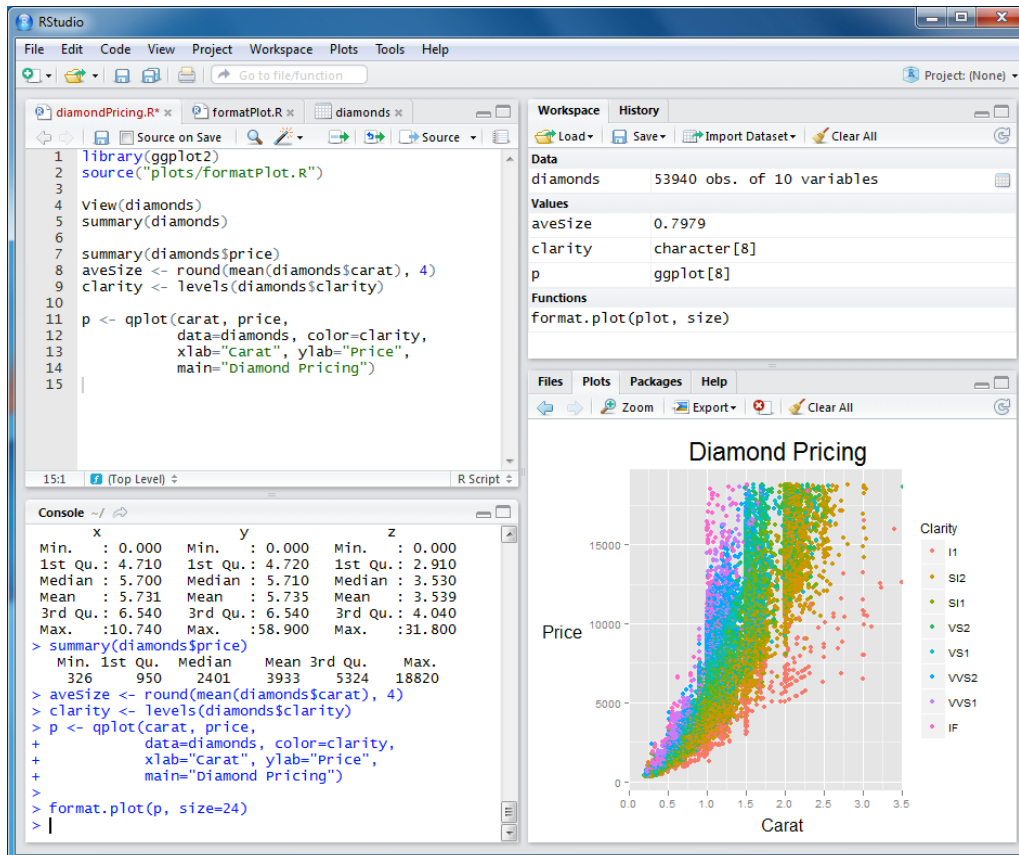
data.table

psych

zoo

janitor

gmodels

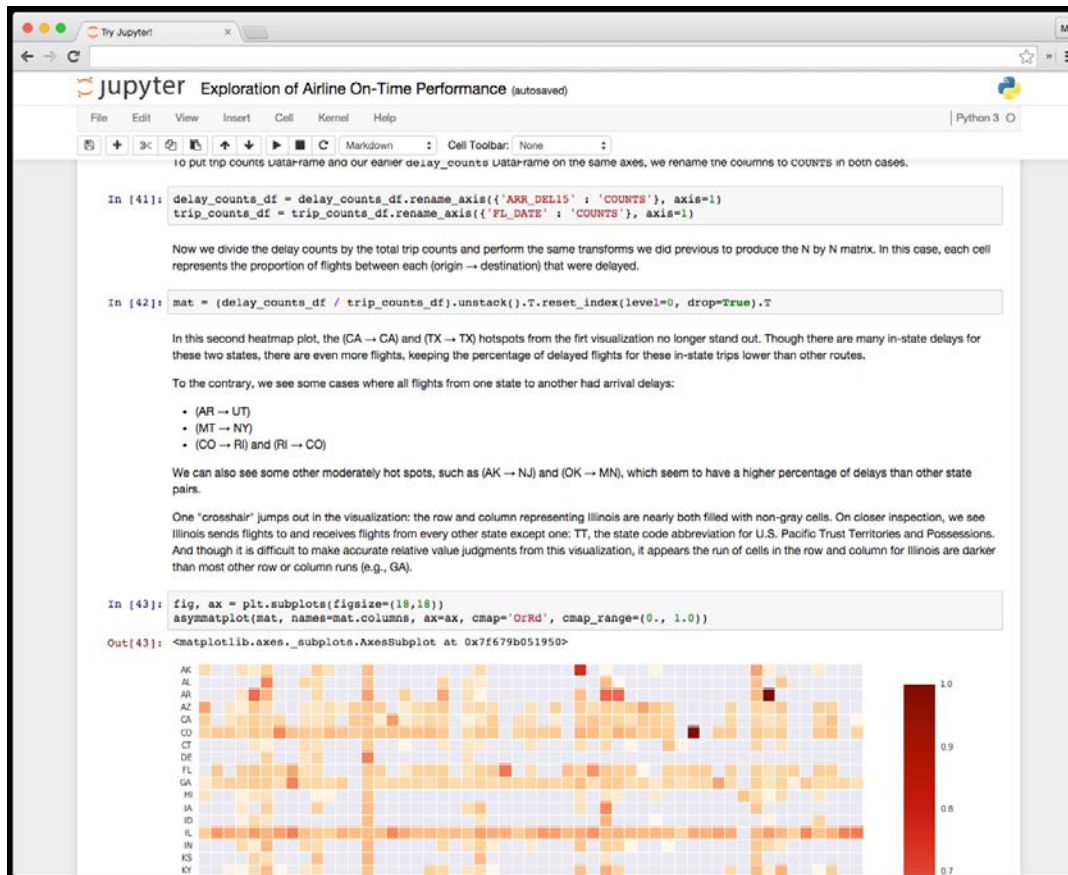




# Python: Free Tool for Coding

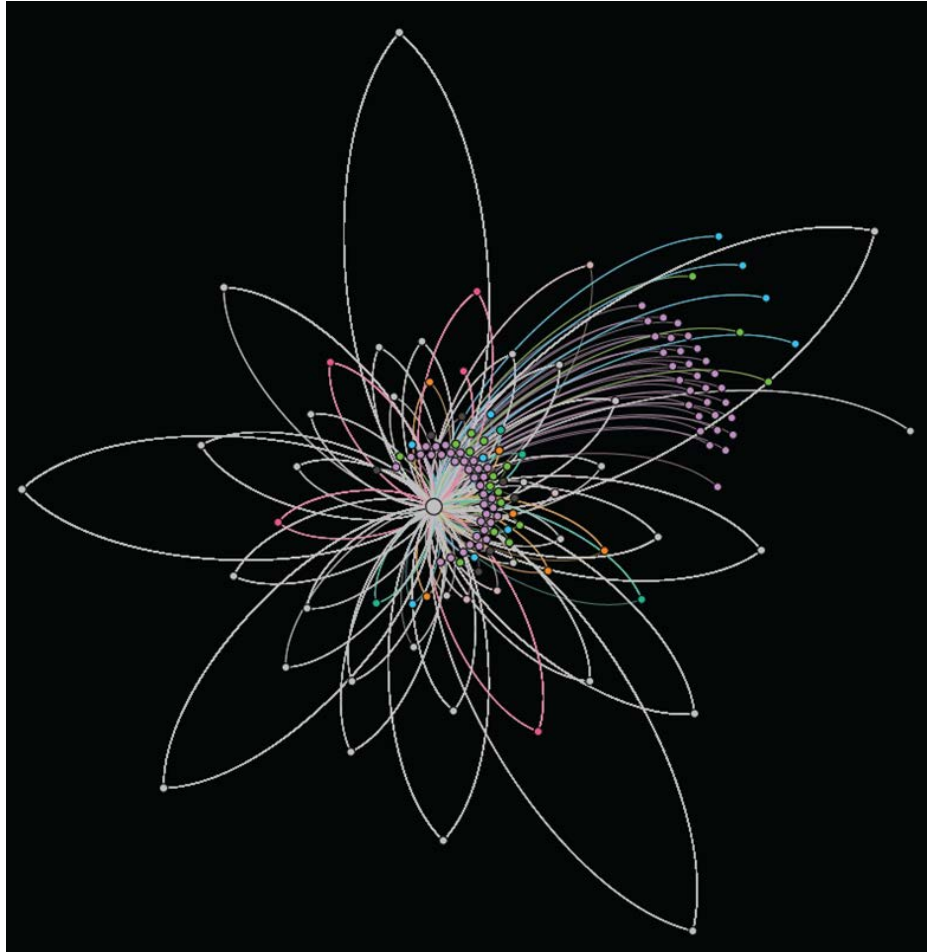


**Packages:**  
pandas  
numpy  
matplotlib

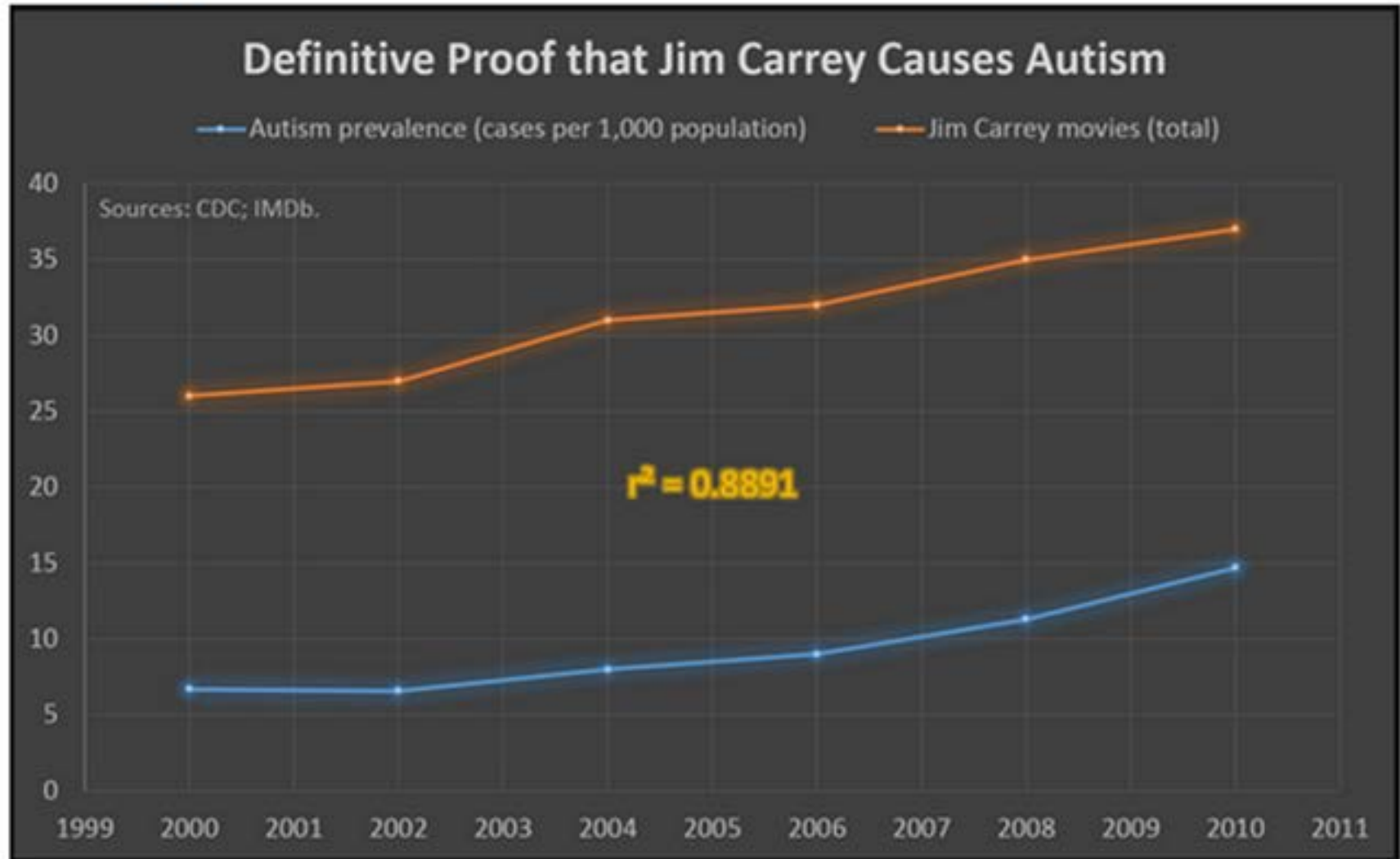




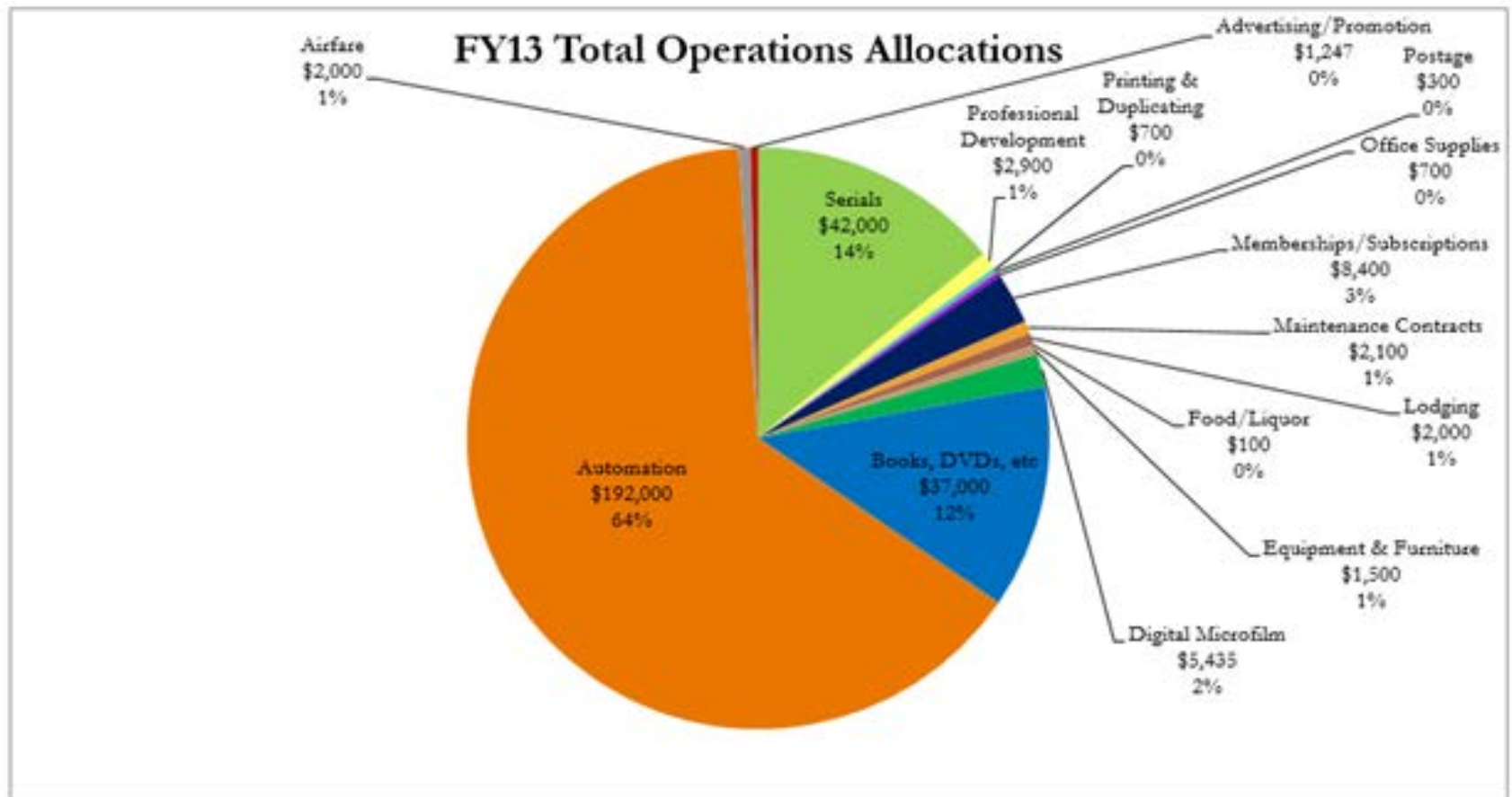
# VISUALIZATION



# #1 Correlation does not equal causation

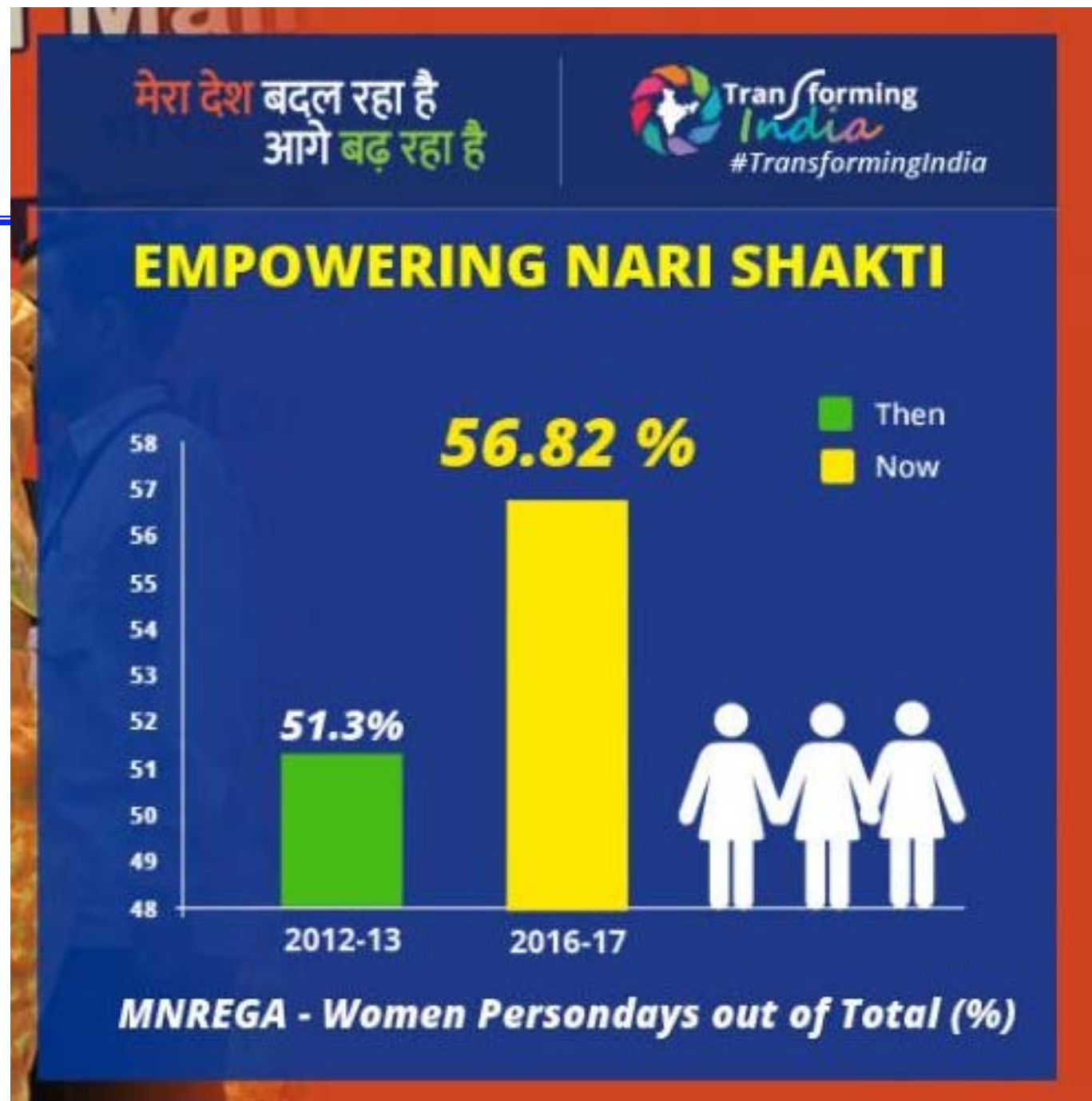


# Rule #2: Put Your Pie Charts on a Diet



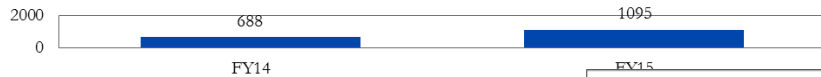
## Rule #3:

Don't  
Mislead



# It's All About the Data

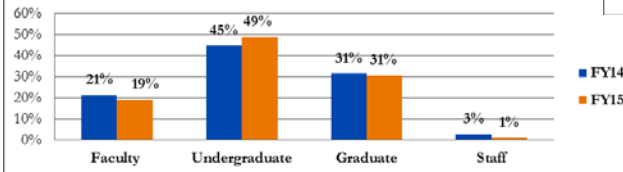
Article\* Requests Comparison...



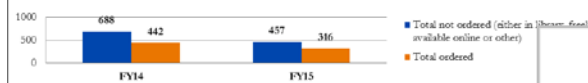
Borrowing Totals FY15 July-December



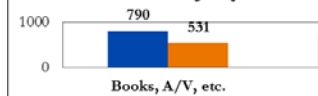
Articles: Comparison of Who Requested in FY14 and FY15 July-December



Articles: Ordered Requests FY14 and FY15 July-December



Overall Requests FY15 July-December



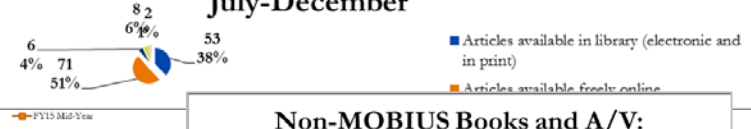
Non-MOBIUS Book and A/V Requests by Rockhurst Patrons by Month FY14 and FY15 July-December



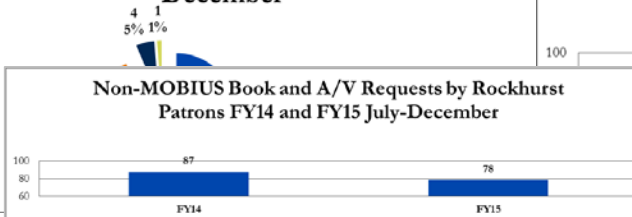
Books and A/V\* Borrowed (OCLC and MOBIUS) FY15 July-December



Articles: Why Requests Were Not Ordered in FY15 July-December



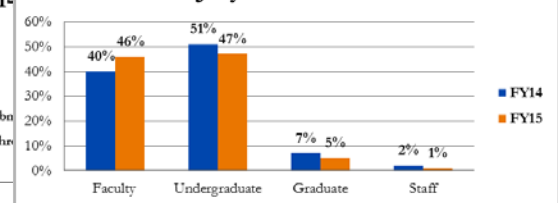
Non-MOBIUS Books and A/V: Who Requested in FY15 July-December



Non-MOBIUS Books and A/V: Ordered Requests FY14 and FY15 July-December



Non-MOBIUS Books and A/V: Who Requested in FY14 and FY15 July-December

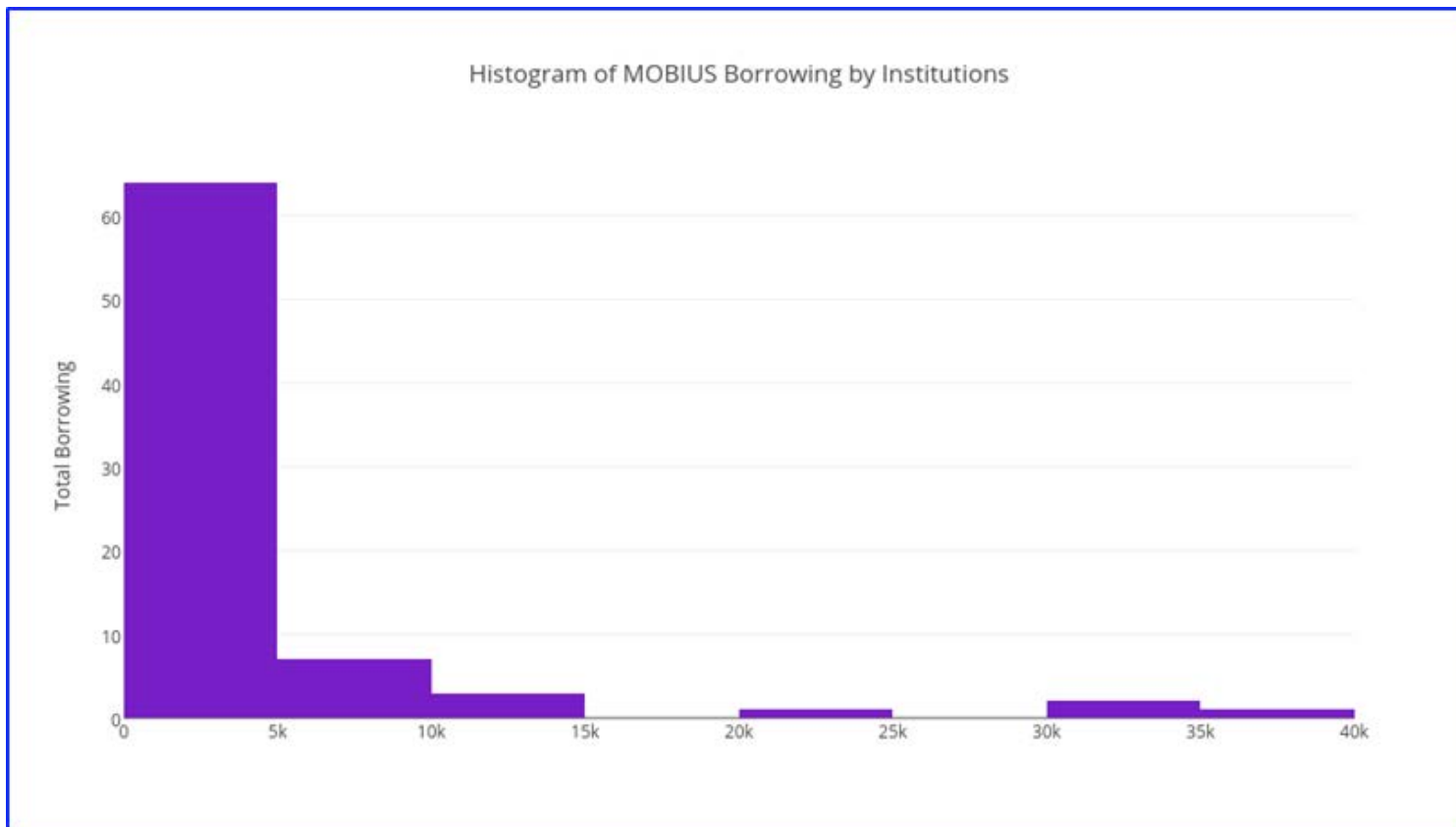




**Make interactive, quality  
graphs online**

- R
- Python (Jupyter)
- D3.js

**....or just make graphs**



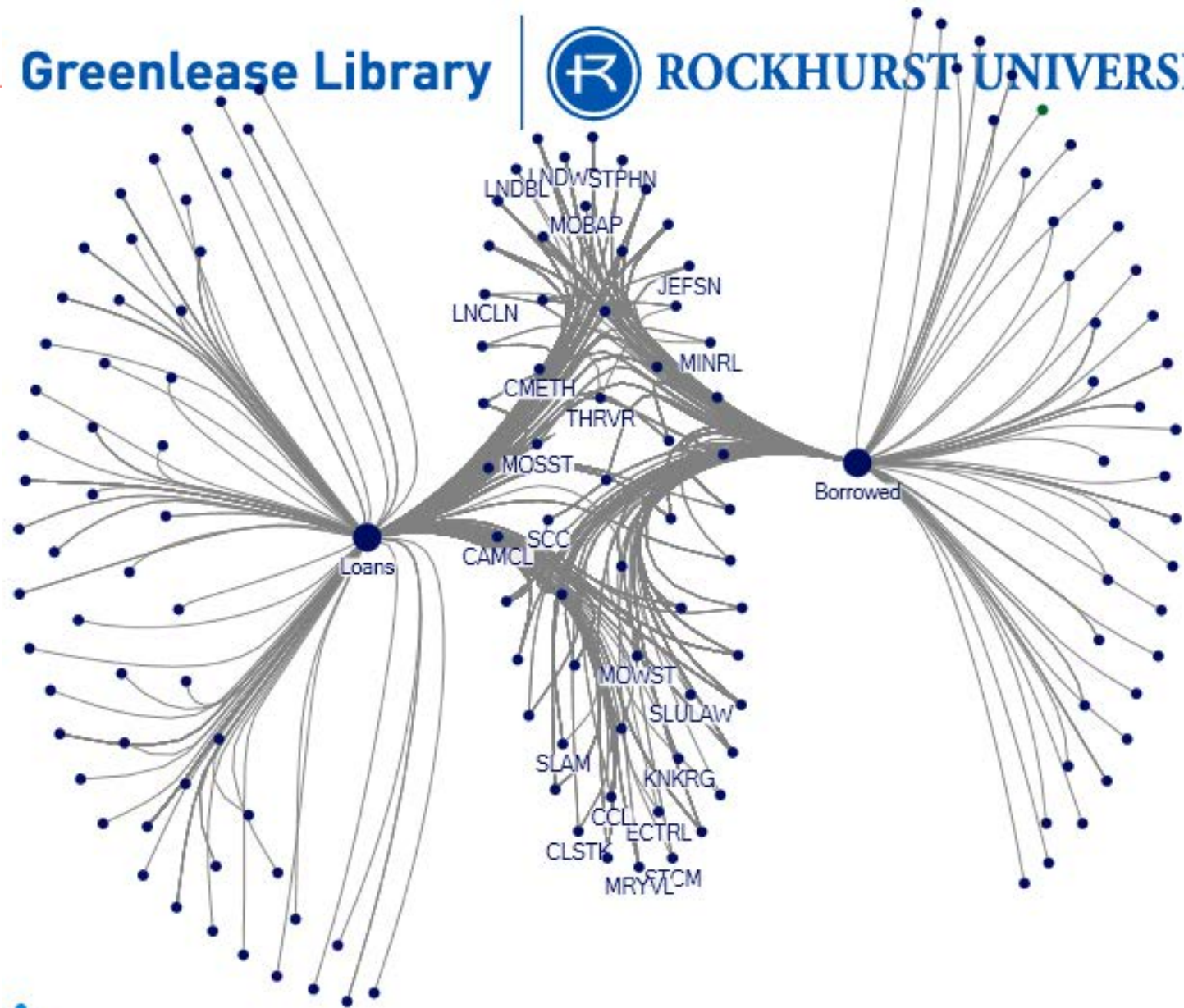
## Story That Made Headlines

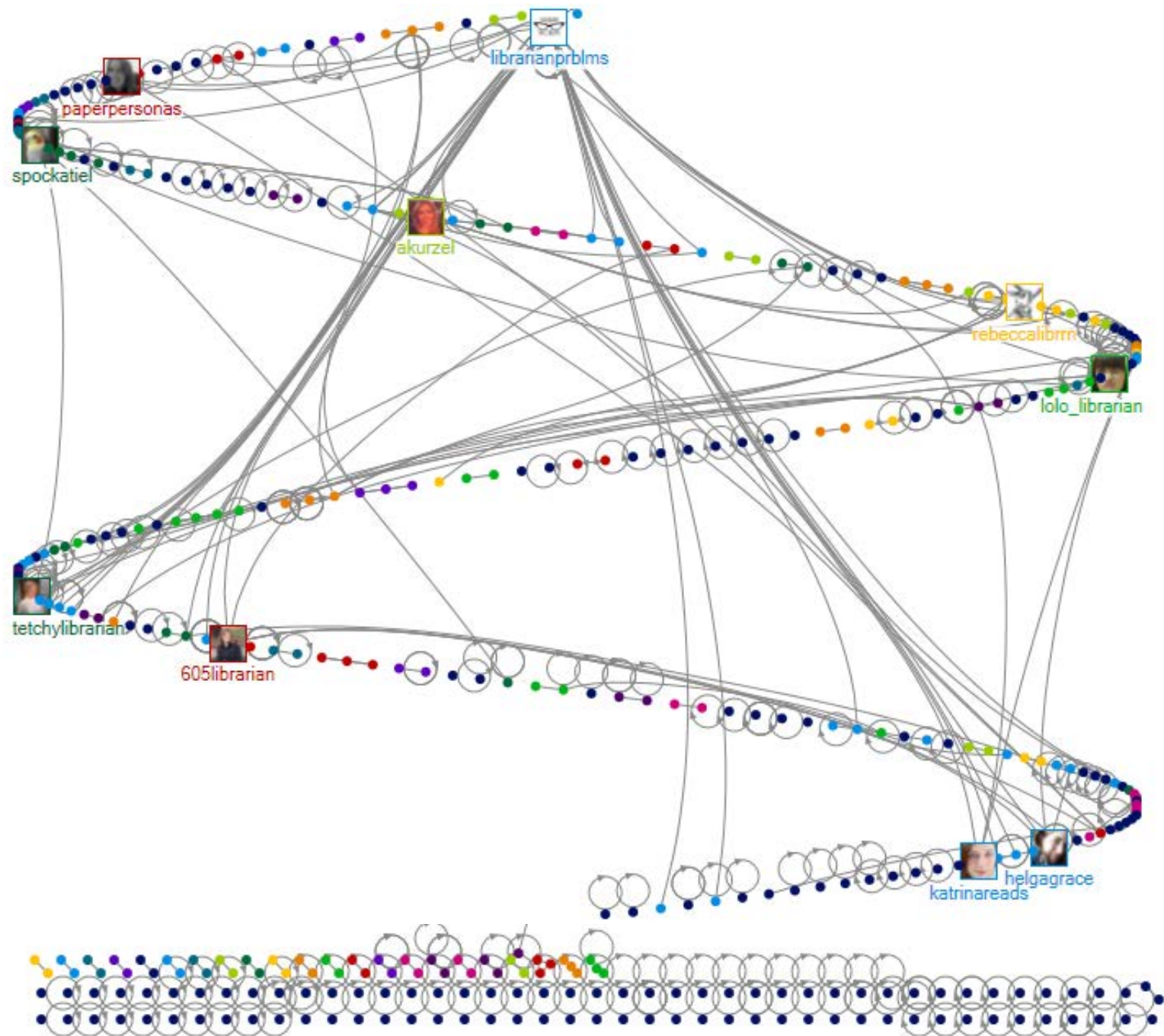


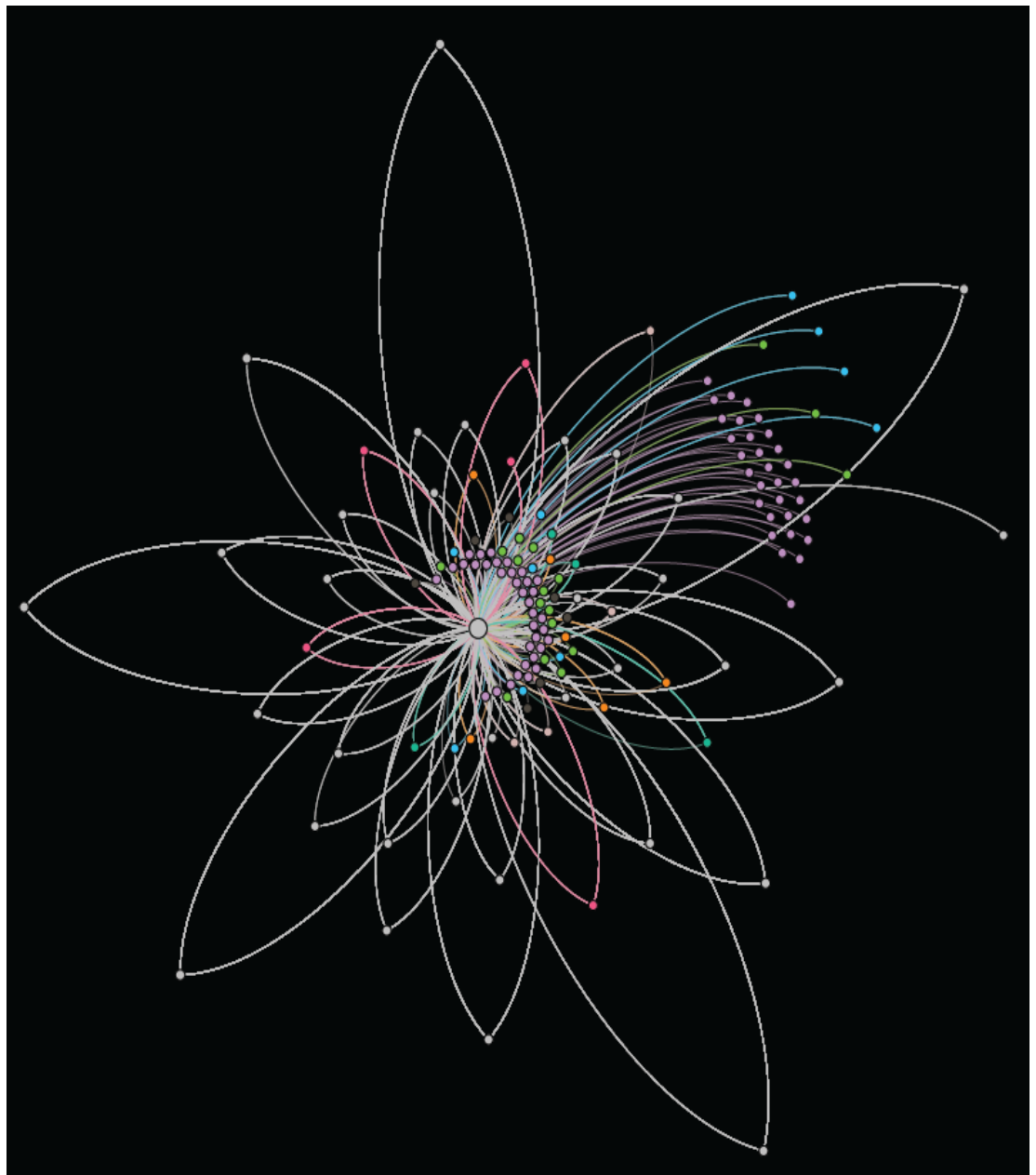
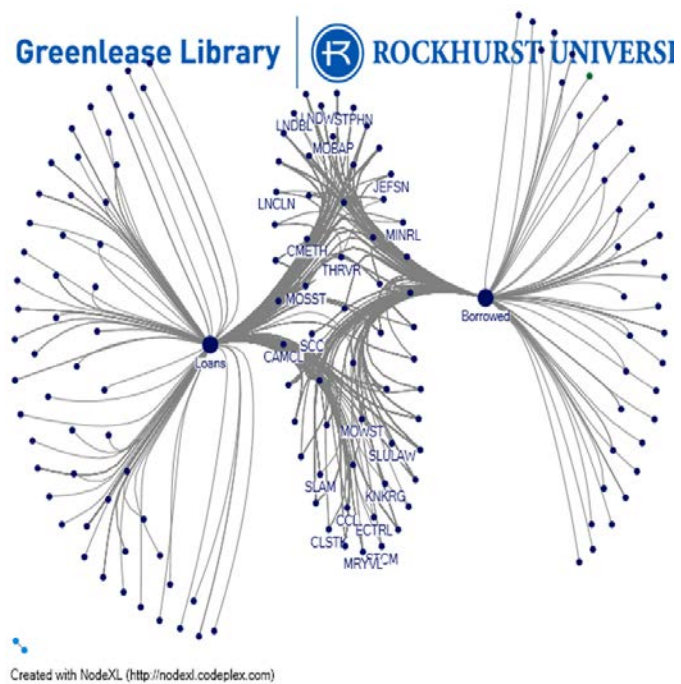
“Numbers have an important story to tell. They rely on you to give them a voice.”











# Tableau

Tableau Software <  
Data visualization company

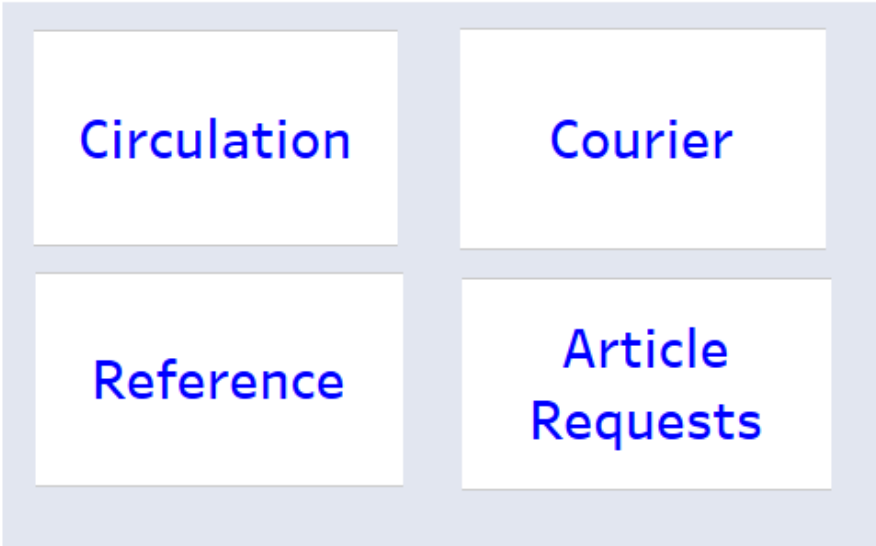


- Data Visualization standard
- Data sharing
- Some analytics
- Public, Educational, Paid
- Steep Learning Curve
- Learn: Coursera, Tableau website, Tableau blog, Youtube videos, Udemy
- <https://public.tableau.com/profile/ellie.kohler/>

Gate Count

286

▼14%



November

Today ▾

16

Classes in Session









Reference Desk Hours



Day at a Glance

10 AM	Flight
1 PM	
7 PM	AS Conf Welcome Reception

Upcoming Student Assistant Absences

October				November
17	18	19	27	7
		 	 	 





November

Today ▾

16

[Classes in Session](#)

## Overdues

Patron Alias	
Anamaria Akerley	0
Antwan Allain	4
Barbera Beauvais	\$ 1
Bernardo Brittingham	39
Beth Bevans	1
Bettie Blakley	2
Caren Coffey	1
Daryl Dougal	8
Dudley Derrico	11
Elfriede Eutsler	5
Fabiola Farrier	\$ 3
Ileen Isabelle	\$ 1
Jannet Janzen	1
Jerrell Jay	1
Joaquin Jaques	1
Jonna Joseph	4
Kayla Khalil	1
Kindra Kasprzak	4
Lan Langevin	3
Lenny Lucious	1
Lidia Lamons	1
Magan McCloud	1
Marcene Marotta	1
Maynard Merlos	1
Myrl Mattern	3
Nyla Niehaus	1
Reyes Rutter	18
Rosalina Reams	1
Salena Seals	1
Suzanne Stokere	2

## Top Checkout Subjects This Month

Reserves

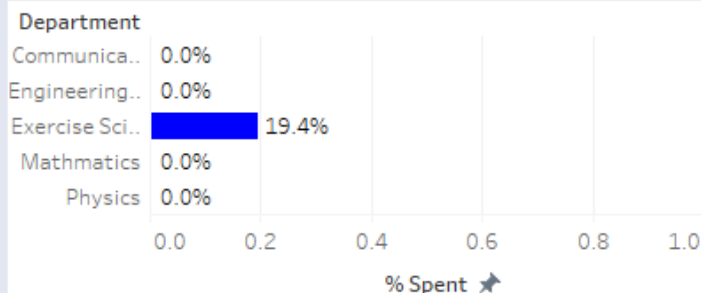
Higher Education

Religion and Theology- Christianity, the Bible, and ..

Literature- Drama

Language- Basque (Uralic)

## Liaison Ordering



Updated Monthly

Ordering Deadline: Saturday March 4, 2017

## Day at a Glance

10 AM Flight

1 PM

7 PM AS Conf Welcome Reception

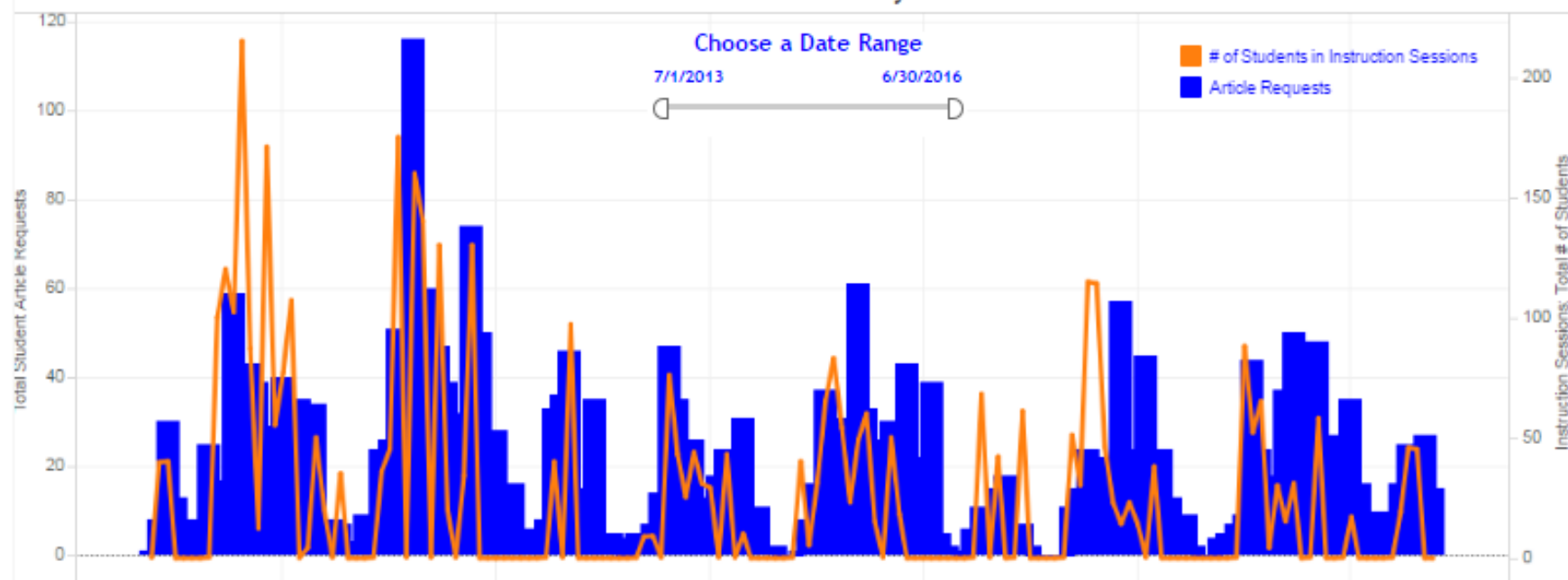
# Instruction and ILL Article Requests: By Student Numbers in Instruction Sessions

Greenlease Library

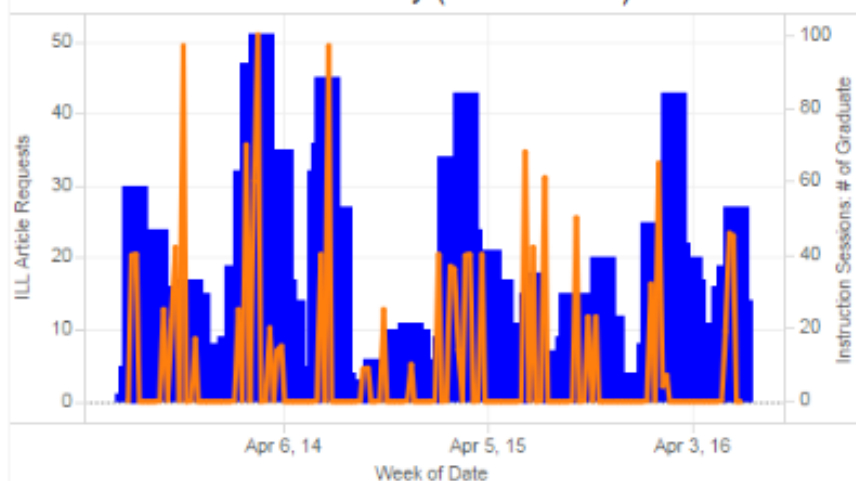


ROCKHURST UNIVERSITY

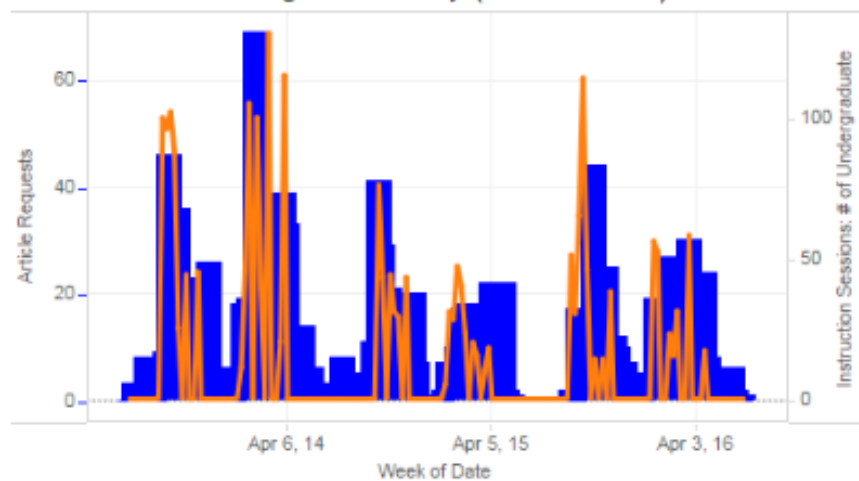
## All Student Activity



## Graduate Only (# of Students)



## Undergraduate Only (# of Students)



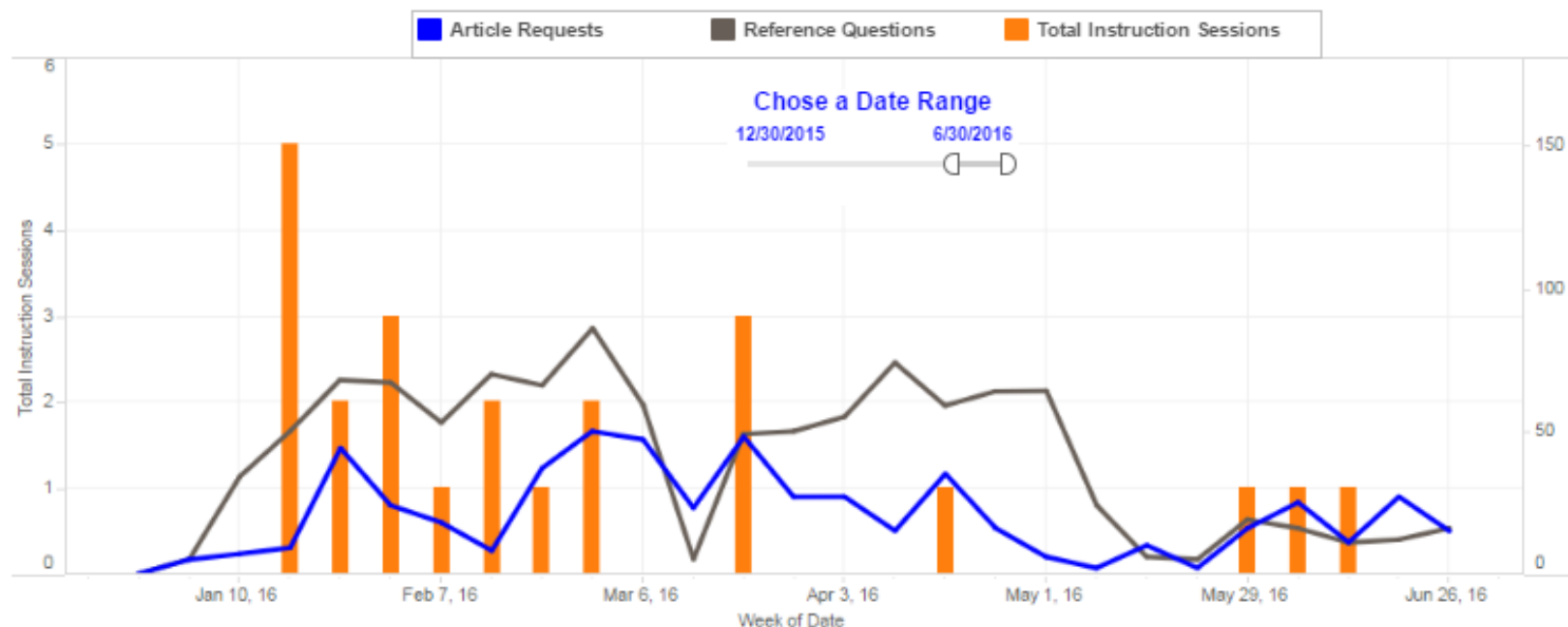


# Instruction, Reference and ILL Article Requests

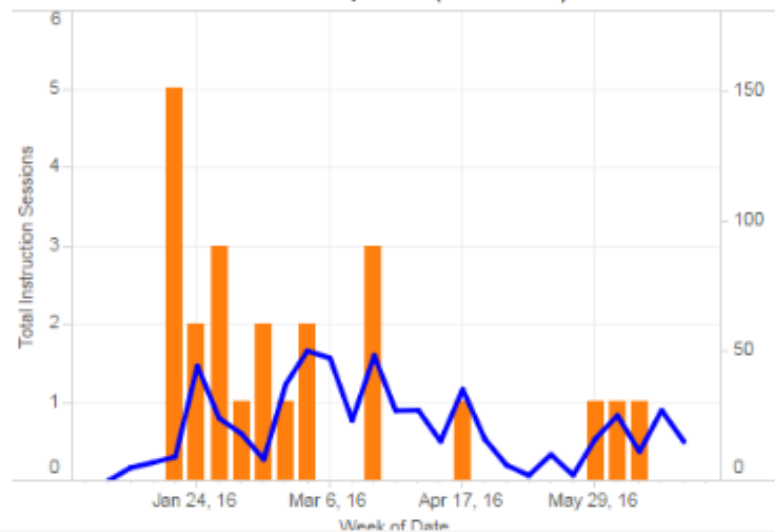
Greenlease Library



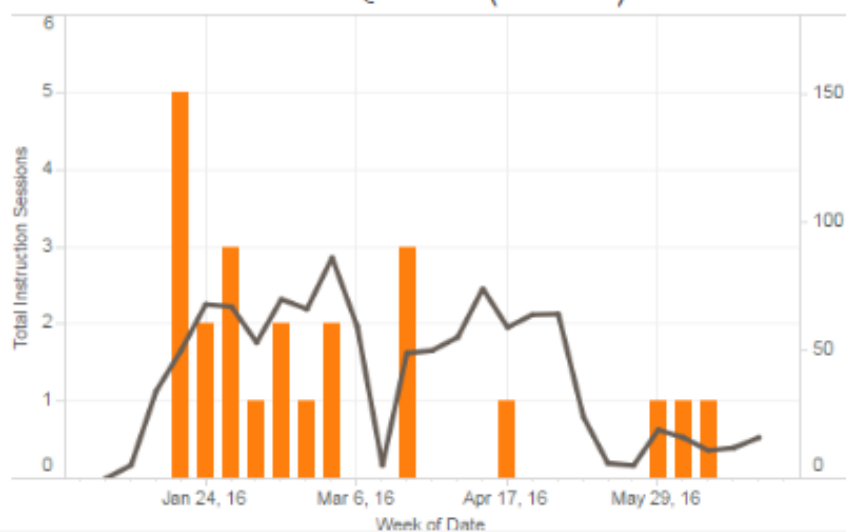
ROCKHURST UNIVERSITY



Instruction and Article Requests (Dual Axis)

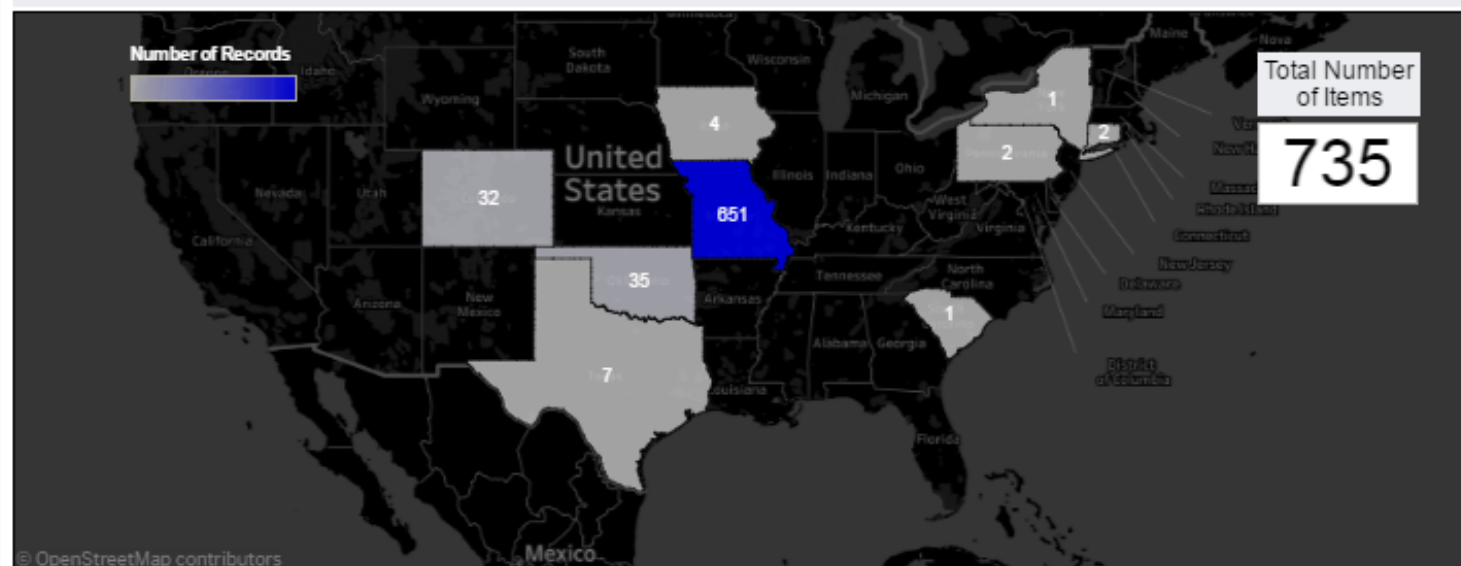


Instruction and Reference Questions (Dual Axis)



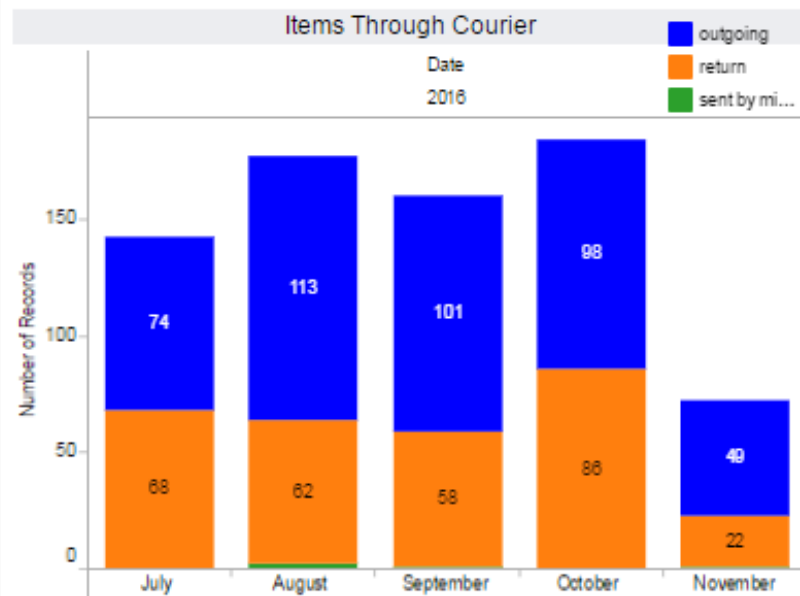


Returnables: Books and AV

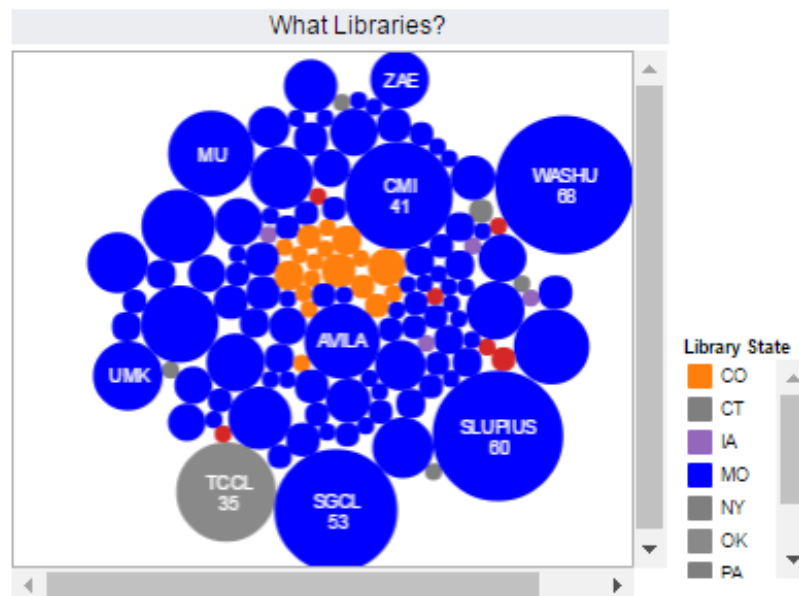


- Status
- ☒ (All)
  - ☒ outgoing
  - ☒ return
  - ☒ sent by mi...

Items Through Courier



What Libraries?



## Observation Study

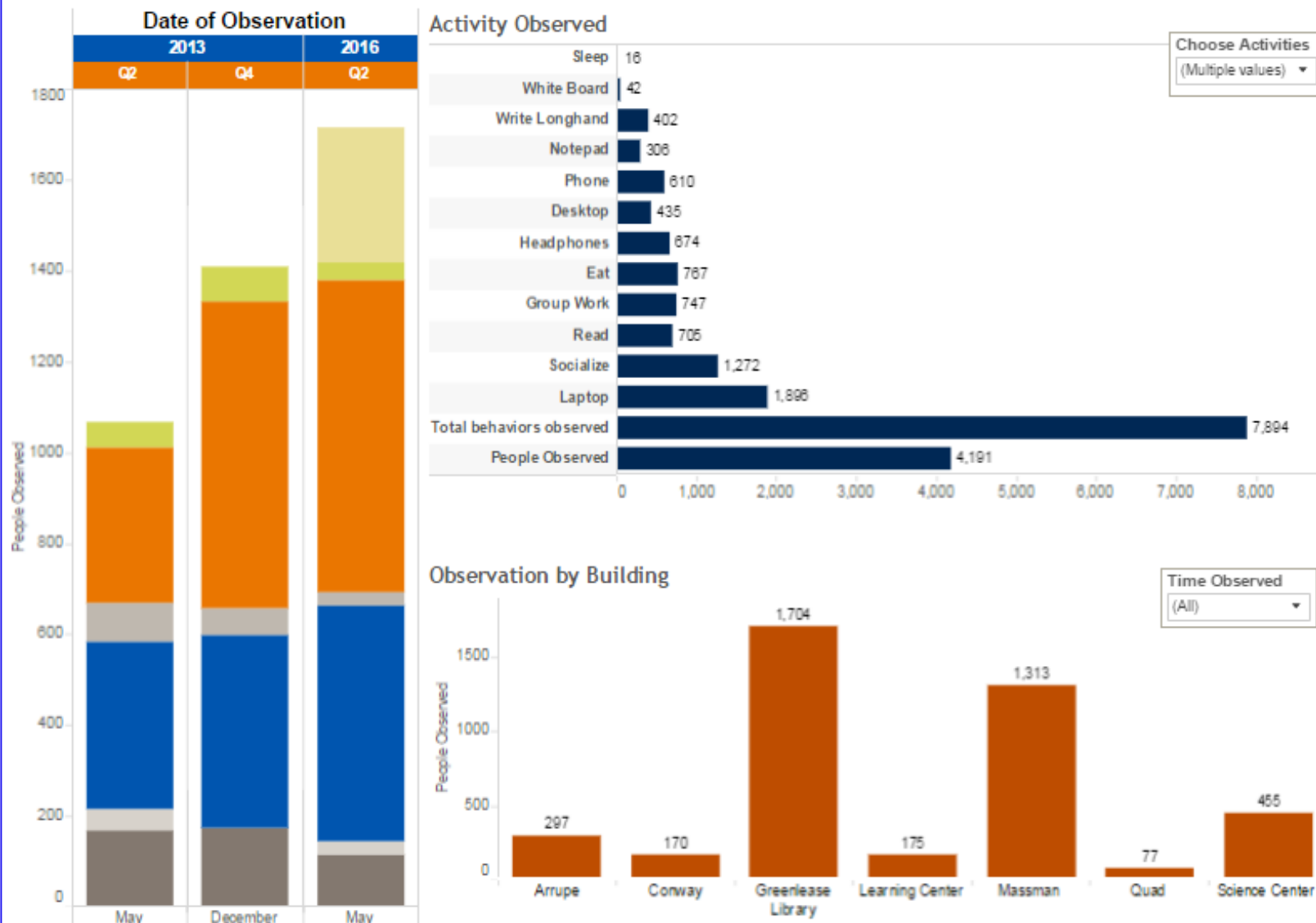
Directions: Click on the Date of Observation or the Building chart to filter the data and see what activities were the most popular.

Tip: You can also filter by Quarter and Year.

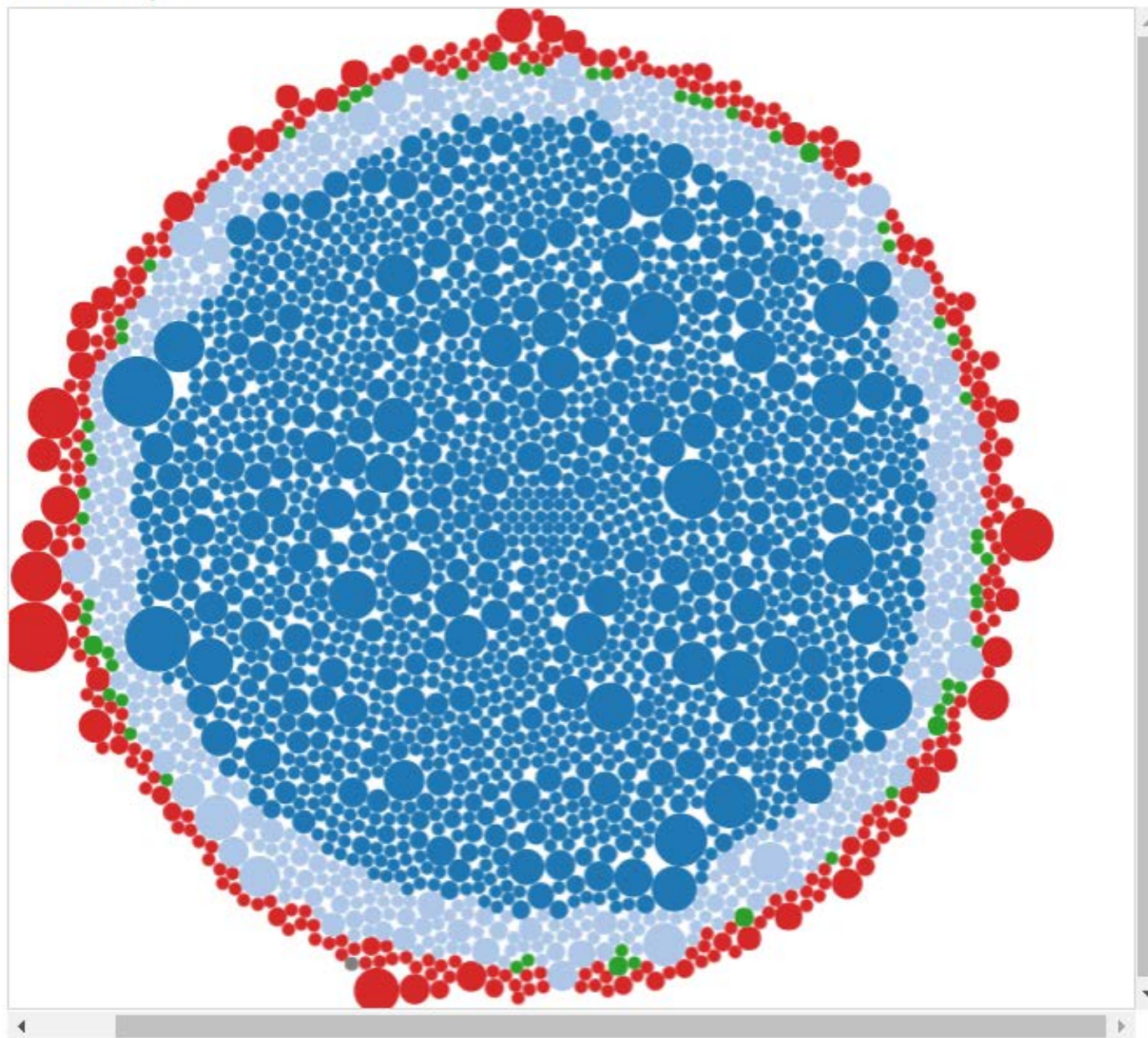


**ROCKHURST  
UNIVERSITY**

*Where leaders learn.*



Journals Titles  
ILL article requests



Status

- ☐ (All)
- ☒ Faculty
- ☒ Graduate
- ☒ Staff
- ☐ Student assistant requ...
- ☒ Undergraduate

Reason not requested

- ☒ Requested
- ☐ Available online
- ☒ Duplicate Request
- ☒ In library database
- ☒ In library print
- ☐ unable to verify citation



---

THANK YOU!

QUESTIONS?

---

[Ellie.Kohler@Rockhurst.edu](mailto:Ellie.Kohler@Rockhurst.edu)