OIRSA | LEHIGH UNIVERSITY

Office of Institutional Research & Strategic Analytics

Effective Reporting and Visualization of Data
Tips and Tricks for Communicating Data

ERAC Staff Development Day 2020

Sumera Ayaz, Business Intelligence Analyst
Caroline Seguin, Business Intelligence Analyst
Office of Institutional Research and Strategic Analytics, Lehigh University

LEHIGH UNIVERSITY | Office of Institutional Research & Strategic Analytics

MY ACCOUNT | LOG OUT

PROVOST'S POLICY | SERVICES | DATA CENTER | DATA GOVERNANCE | RELATED LINKS | CONTACT US

Data Center

Home » Data Center

- Provost's Policy
- Services
- Data Center
  - Data Visualization Tools
  - Census Data
  - Common Data Set

The OIRSA Data Center is home to current and past information about Lehigh University, including Census Data, the Common Data Set, the University Profile, our new visualization tools and a variety of other Reports & Presentations.

And, although many university constituents may not realize it, the data we compile each year appear throughout the university in a variety of familiar formats, including Lehigh at a Glance and the Undergraduate Admissions Class

LEHIGH UNIVERSITY | Office of Institutional Research & Strategic Analytics

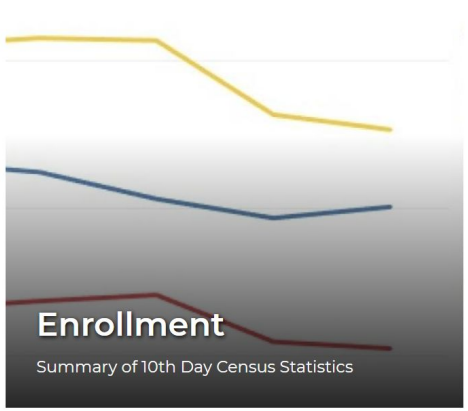
PROVOST'S POLICY | SERVICES | DATA CENTER | DATA GOVERNANCE | RELATED LINKS | CONTACT US

Data Visualization Tools

Home » Data Center » Data Visualization Tools

- Provost's Policy
- Services
- Data Center
  - Data Visualization Tools
    - Enrollment
    - Admissions
    - Graduation & Retention
    - Faculty & Staff
  - Census Data
  - Common Data Set
  - University Profile

View the latest data visualizations available for each of the subject areas below.



Agenda

Part 1: Tips and Tricks

1. Who and what is the report for?
2. Purposefully simplify the visualizations
3. Provide context for the data and the information
4. Iterative process

Part 2: Examples

Tips and Tricks

1. Who and what is the report for?

Who will use the report?

- Who requested this report?
- Do they have permission to see the data?
  - Individual records vs Aggregated view?
  - FERPA
    - https://lts.lehigh.edu/services/explanation/ferpa-lehigh
    - https://ras.lehigh.edu/content/ferpa-faculty
  - Restrict access to a single user/office?
  - User specific access?
- Who will be using it to make decisions/to answer questions?
- How familiar are they with the data?
- Is it possible that the report will be shared outside of the requester group? (Other offices? Outside of Lehigh?)

What is the goal of the report?

- What information is the client requesting? What decision(s) is it supporting? What is its intended use?
- What key findings are you trying to communicate?
- Do they need the report as part of a bigger process?
- Report style?
- Do you need to visualize the data? Visually representing your data can help users process complex information much easier.
- When should the report be delivered? Report frequency?

At the end of this step you should...

- Know what metrics you should include in your report
- Know how you want to slice and dice the data
- Have an idea of the amount of details you will need to provide
- Have a first idea of what graphics you would want to include
  - The type of graphic
  - The number of graphics you might want to include
- Know what reporting/visualization software you would use
- Report frequency

Categories of reports

Ad hoc reports - Ad hoc request (not recurring), quick turnover.Often created in Excel.

Transactional reports - Recurring request, where the purpose is to extract operational data (i.e. Banner) but not for analysis purposes.Often created in Argos.

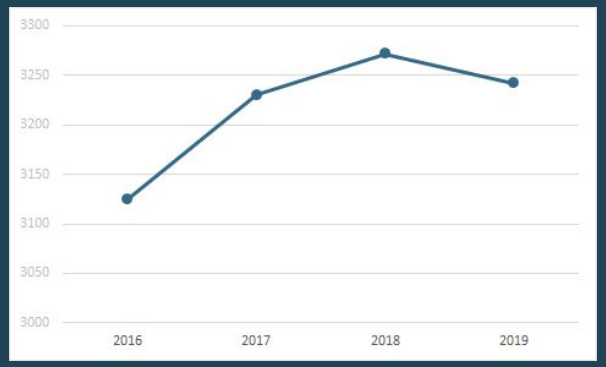
Analytical reports - Recurring request, where the purpose is for analysis of the data, and where visualizations need to be created. Often created in Tableau, Excel.

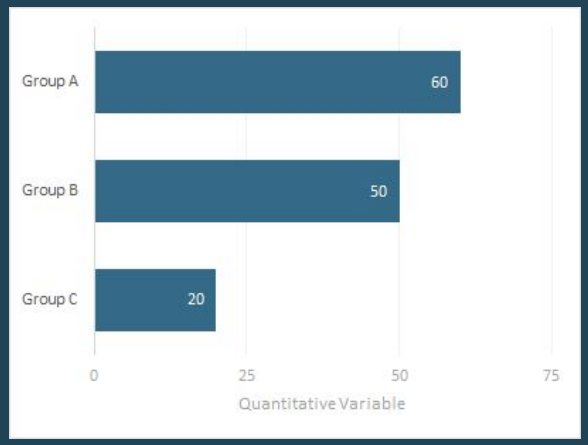
2. Purposefully simplify the visualizations

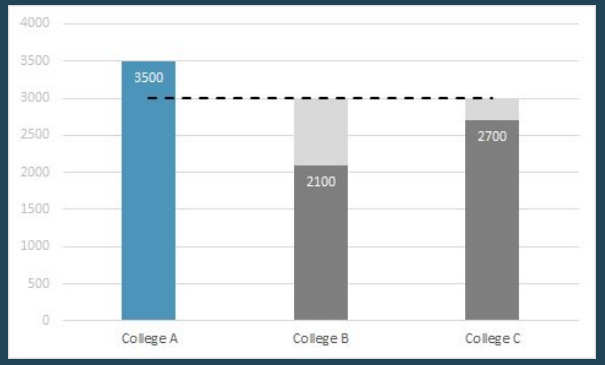
Choose a chart type

Choose a chart type that serves what you are communicating:

- Change in time (increase/decrease)?
- Comparing a metric across categories?
- Progress towards a goal/target?







Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith, Chris Campbell, Ian Bott, Liz Faunce, Graham Parrish, Billy Ehrenberg-Shannon, Paul McCallum, Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

ft.com/vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit; climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar

Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

Spine

Splits a single value into two contrasting components (eg male/female).

Surplus/deficit filled line

The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment; income and life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline

A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot

Usually used to show how the relationship between 2 variables has changed over time.

Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar

Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column

See above.

Ordered proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot

Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Lollipop

Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram

The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

Dot plot

A simple way of showing the change or range (min/max) of data across multiple categories.

Dot strip plot

Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Boxplot

Summarise multiple distributions by showing the median (centre) and range of the data

Violin plot

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

Population pyramid

A standard way for showing the age and sex breakdown of a population distribution, effectively, back to back histograms.

Cumulative curve

A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasise individual points in a distribution. Points can be sized to an additional variable. Best with medium-sized datasets.

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries. Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column

Columns work well for showing change over time - but usually best with only one series of data at a time.

Column + line timeline

A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope

Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart

Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Candlestick

Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projections)

Use to show the uncertainty in future projections - usually this grows the further forward to projection.

Connected scatterplot

A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

Calendar heatmap

A great way of showing temporal patterns (daily, weekly, monthly) – at the expense of showing precision in quantity.

Priestley timeline

Great when date and duration are key elements of the story in the data.

Circle timeline

Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).

Vertical timeline

Presents time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram

Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph

A type of area chart; use when seeing changes in proportions over time is more important than individual values

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation, volumes in general

Column

The standard way to compare the size of things. Must always start at 0 on the axis.

Bar

See above. Good when the data are not time series and labels have long category names.

Paired column

As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Paired bar

See above.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Isotype (pictogram)

Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

Lollipop

Lollipop charts draw more attention to the data value than standard bar/column – does not have to start at zero (but preferable).

Radar

A space-efficient way of showing value of multiple variables – but make sure they are organised in a way that makes sense to reader.

Parallel coordinates

An alternative to radar charts – again, the arrangement of the variables is important. Usually benefits from highlighting values.

Bullet

Good for showing a measurement against the context of a target or performance range.

Grouped symbol

An alternative to bar/column charts when being able to count data or highlight individual elements is useful.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Fiscal budgets, company structures, national election results

Stacked column/bar

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Pie

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut

Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).

Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Voronoi

A way of turning points into areas – any point within each area is closer to the central point than any other centroid.

Arc

A hemicycle, often used for visualising parliamentary composition by number of seats.

Gridplot

Good for showing % information, they work best when used on whole numbers and work well in small multiple layout form.

Venn

Generally only used for schematic representation.

Waterfall

Can be useful for showing part-to-whole relationships where some of the components are negative.

Spatial

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rate/ratio)

The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Proportional symbol (count/magnitude)

Use for totals rather than rates – be wary that small differences in data will be hard to see.

Flow map

For showing unambiguous movement across a map.

Contour map

For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values

Equalised cartogram

Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

Scaled cartogram (value)

Stretching and shrinking a map so that each area is sized according to a particular value.

Dot density

Used to show the location of individual events/locations – make sure to annotate any patterns the reader should see.

Heat map

Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information; relationship graphs.

Sankey

Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord

A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

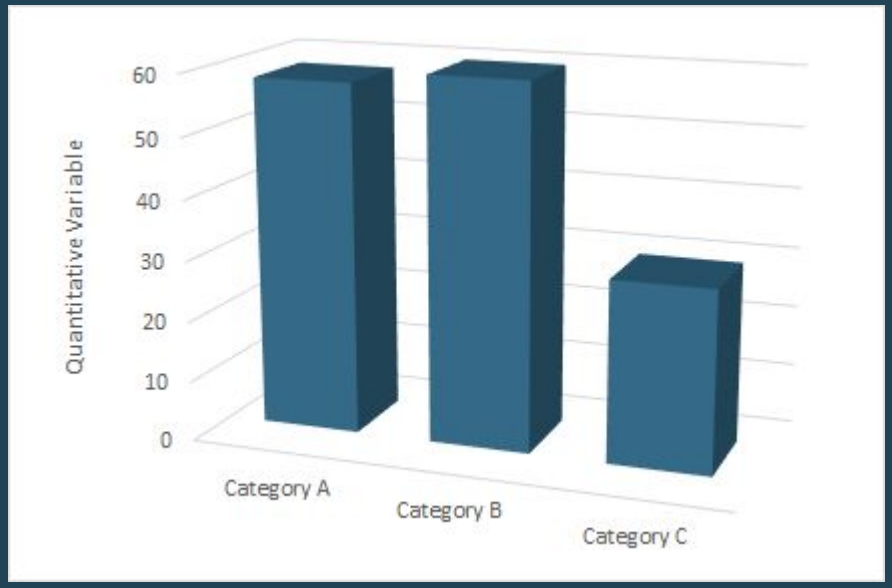
Network

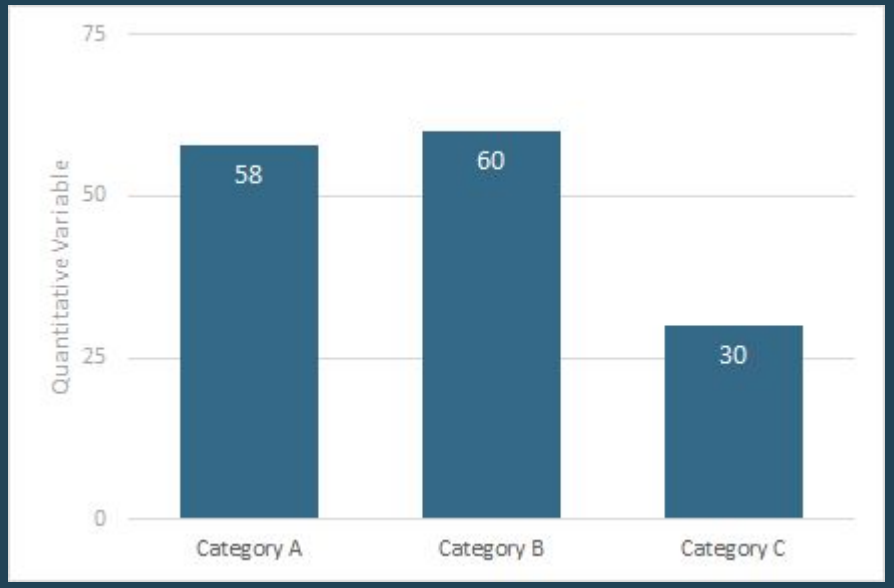
Used for showing the strength and inter-connectedness of relationships of varying types.

© Financial Times 2016–2019
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Simplify your visualizations

- Remove any 3D effect

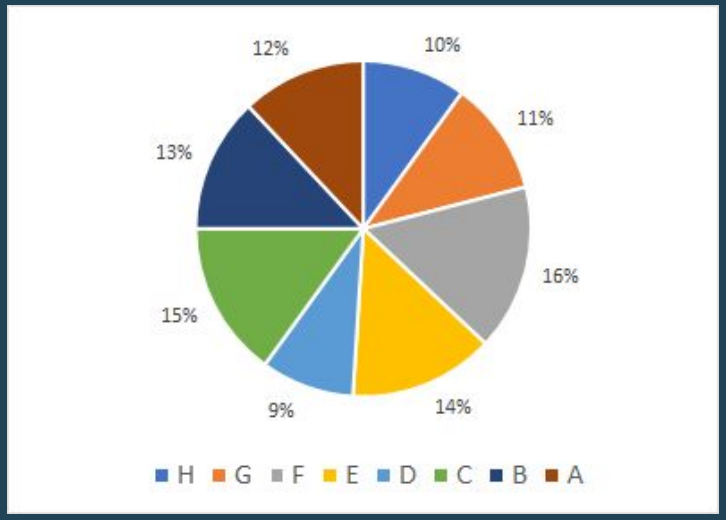


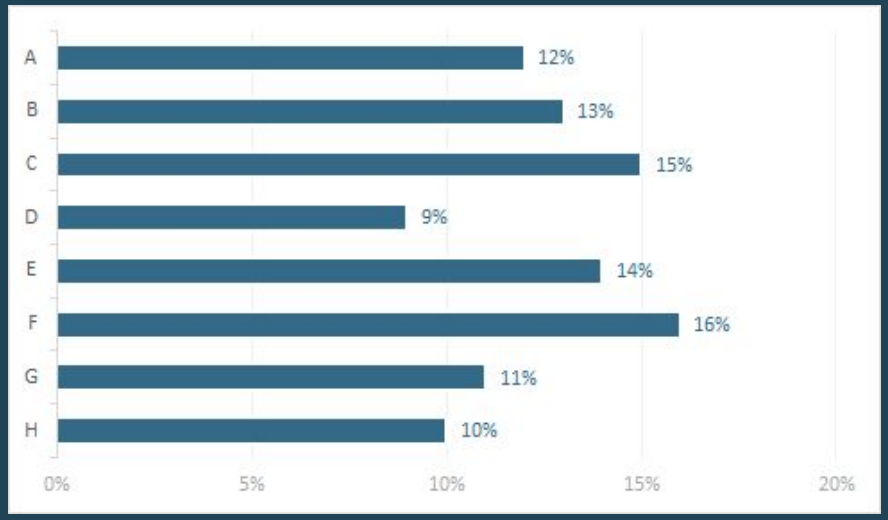


Simplify your visualizations

Make sure the graph type is as easy to read as possible:

- Remove any 3D effect
- Consider alternatives to pie charts

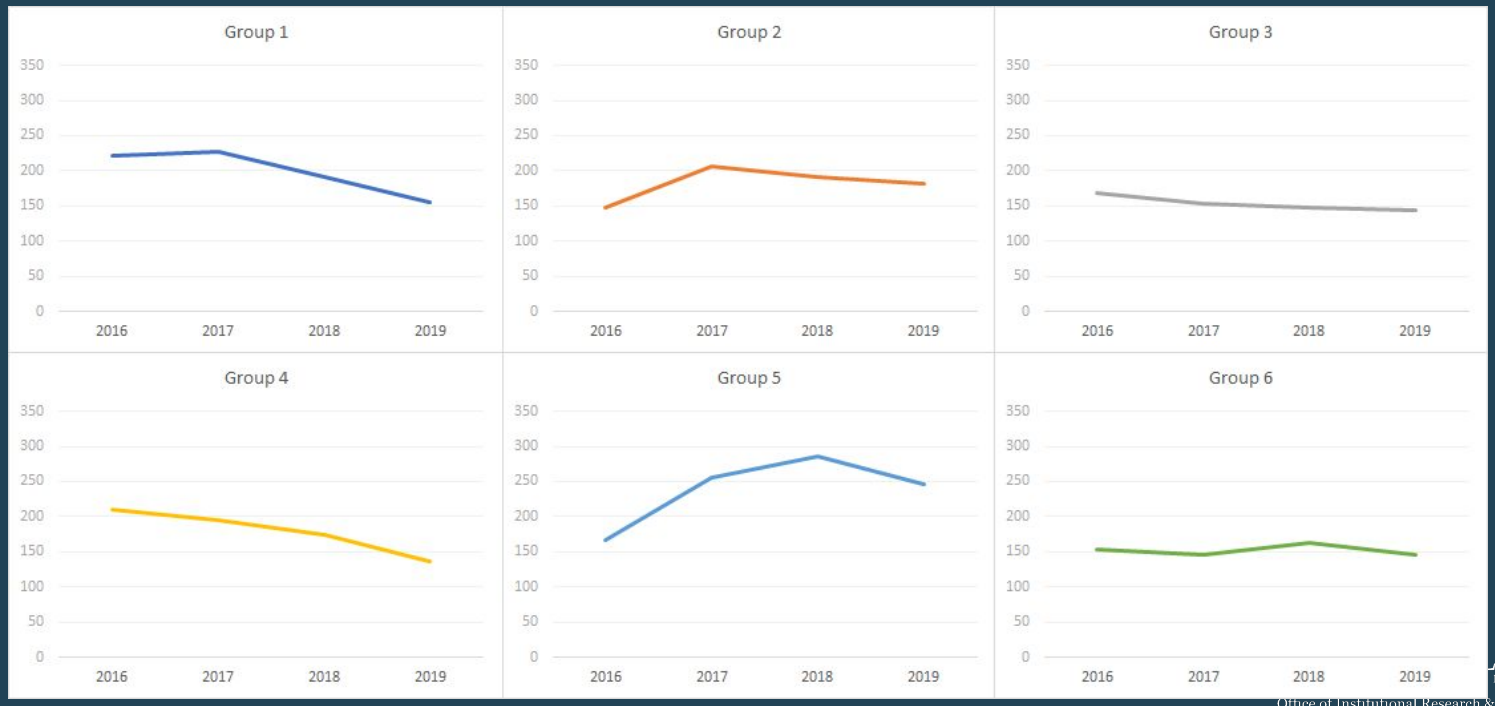




Simplify your visualizations

Make sure the graph type is as easy to read as possible:

- Remove any 3D effect
- Consider alternatives to pie charts
- Is the graph going to be too crowded?



Refine your visualizations

*1- Use **color** with intention:*

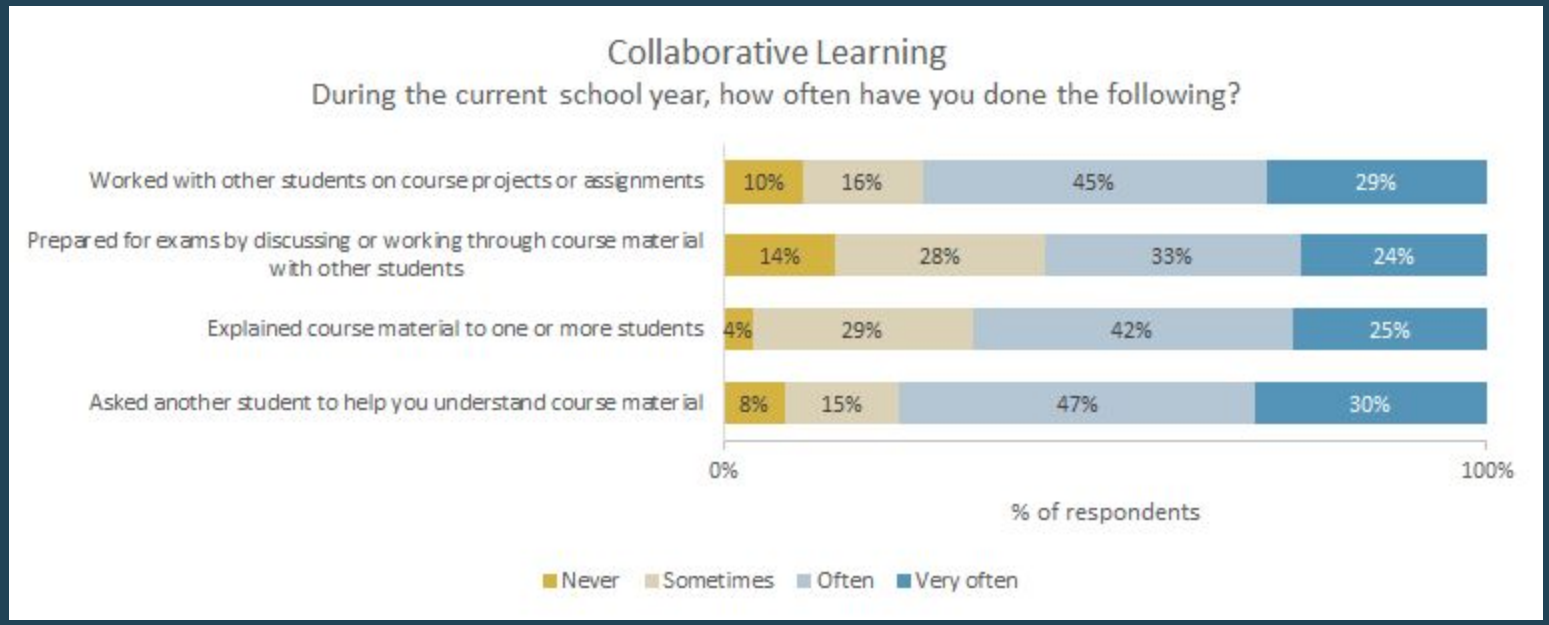
- Use color sparingly
- Purposeful color choices: move beyond software defaults
- Same color for same data field
- Keep in mind color blindness

*2- Visually **highlight** the key findings:* use annotations and color to your advantage.

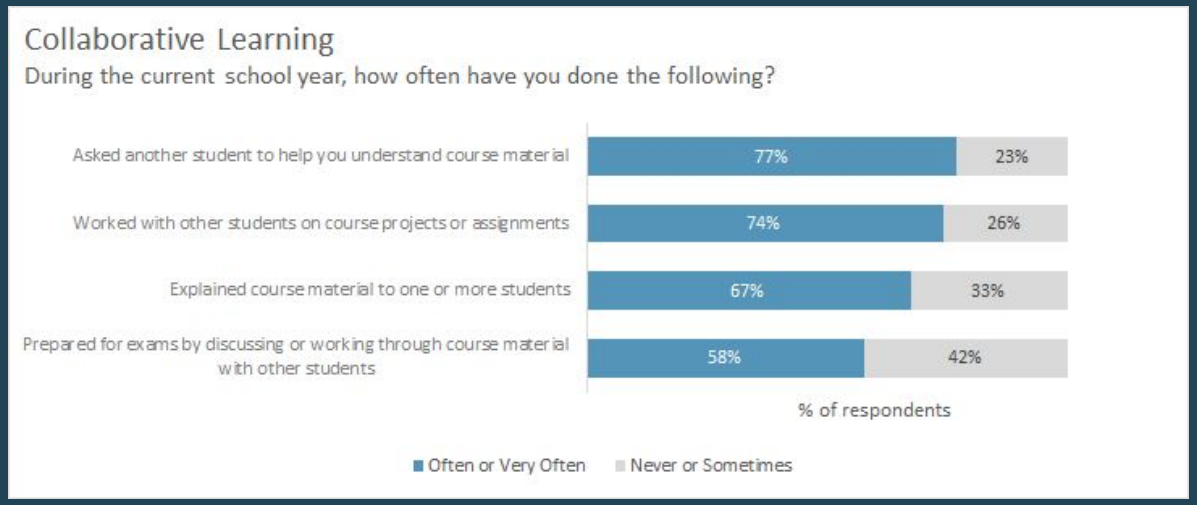
3- Further improvements:

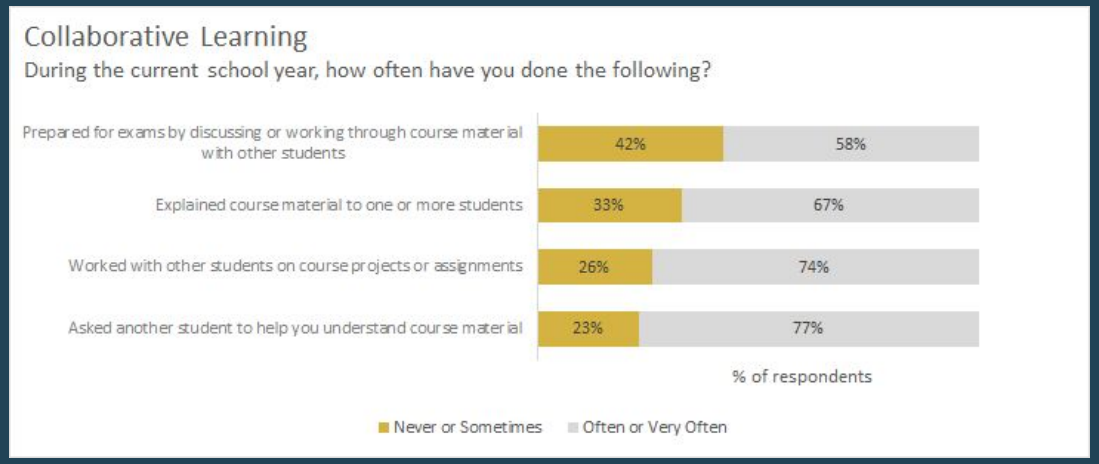
- Direct labeling when possible
- Order the data in a way that serves the graph
- Remove visual noise: https://images.squarespace-cdn.com/content/56713bf4dc5cb41142f28d1f/1450306653111-70K5IT30R69NWPDIE1ZJ/data-ink.gif?content-type=image%2Fgif (credit: darkhorseanalytics.com)

Refine your visualization - Examples



Refine your visualization - Examples





Refining the report: Branding

- Lehigh branding
  - https://www1.lehigh.edu/communications/assets-resources
  - https://www.lehigh.edu/~inis/sdownloads/Lehigh-Brand-Visual-Guide.pdf
- OIRSA branding
- Other offices branding

3. Provide context for the data and the information

Provide context for the information

- For all graphics (including tables)
  - Descriptive title
  - Descriptive subtitle
  - Descriptive axis labels
- Add text to add useful information about what is being represented
- Summarised descriptions help users better understand the report

Provide context for the data

- List the data source(s). Where was the data extracted from?
- Date of data extraction/collection?
- Provide data definitions where necessary
  - Consult Lehigh's Data Cookbook: https://lehigh.datacookbook.com/

4. Iterative Process

Clean your report

Structure:

- Tell a story with the data
- Consider the visual hierarchy of the report

Visual design:

- Reduce the information on the page
- Use blank space wisely
- Remove visual noise
- Alignment of text/graphs

Get feedback on the report

- Have a coworker test out the report:
  - How quickly do they understand what is being represented, and the key findings?
  - If the report is interactive, how clear is this interaction?
  - Are they being naturally guided through the report?
  - What suggestions would they have to make the report better, and the data clearer?

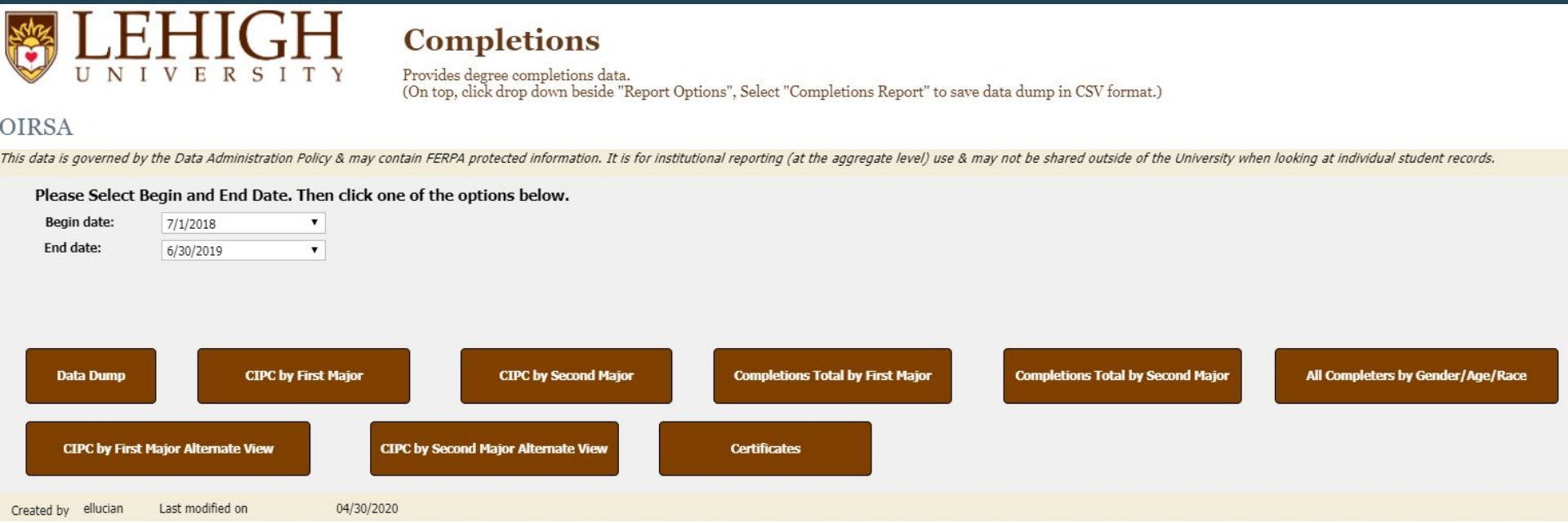
- Get feedback from the requester

Examples

Examples

1. Completions - Argos report
2. Lehigh University Art Gallery - Excel report
3. Lehigh University Art Gallery - Tableau report

1. Completions - Argos Report



LEHIGH UNIVERSITY

Completions

Return to Main Menu

Provides degree completions data.
(On top, click drop down beside "Report Options", Select "Completions Report" to save data dump in CSV format.)

OIRSA

This data is governed by the Data Administration Policy & may contain FERPA protected information. It is for institutional reporting(at the aggregate level) use & may not be shared outside of the University when looking at individual student records.

By Gender

Measures | Dimensions | OLAP Saved Settings

DEGREE TYPE	Bachelors	Certification	Doctorate	Masters	Tot
GENDER	PIDM COUNT	PIDM COUNT	PIDM COUNT	PIDM COUNT	PIDM
	Value	Value	Value	Value	Value
F	522	10	43	236	
M	624	11	64	302	
Total by Columns	1146	21	107	538	

By Age

Measures | Dimensions | OLAP Saved Settings

DEGREE TYPE	Bachelors	Certification	Doctorate	Masters	Total by
AGE GROUP	PIDM COUNT	PIDM COUNT	PIDM COUNT	PIDM COUNT	PIDM COUN
	Value	Value	Value	Value	Value
[Null]	0	0	2	8	
18-24	1123	17	1	250	
25-39	23	3	94	247	
40 and above	0	1	10	33	
Total by Columns	1146	21	107	538	

By Race

Measures | Dimensions | OLAP Saved Settings

DEGREE TYPE	Bachelors	Certification	Doctorate	Masters	Total by Rows
REPORT ETHINC...	PIDM COUNT	PIDM COUNT	PIDM COUNT	PIDM COUNT	PIDM COUNT
	Value	Value	Value	Value	Value
American Indian or Alaskan	1	0	0	0	1
Asian	106	1	0	15	121
Black or African American	33	0	1	15	49
Hispanic	108	0	1	28	136
Native Hawaiian and Other	0	0	0	1	1
Non Resident Alien	92	1	53	192	337
Race/Ethnicity unkown					

Created by sua216 Last modified on 11/20/2019

2. Lehigh University Art Gallery - Excel report

Student Aggregates for Q1

Undergraduate Level	
College	**Count**
Arts Engineering	6
Arts & Sciences	146
Business	55
P.C.Rossin Engrg & Applied Sci	79
Intercollegiate Programs	35
Grand Total	**321**

Graduate Level	
College	**Count**
Arts & Sciences	3
Business	3
Education	4
P.C.Rossin Engrg & Applied Sci	18
Grand Total	**28**

Breakdown by expected graduation date

Undergraduate Level		
College	**Exp Graduation Date**	**Count**
Arts Engineering	18-MAY-20	1
Arts Engineering	22-MAY-23	4
Arts Engineering	23-MAY-22	1
Arts Engineering Total		**6**
Arts & Sciences	18-MAY-20	19
Arts & Sciences	22-MAY-23	73
Arts & Sciences	23-MAY-22	31
Arts & Sciences	24-MAY-21	23
Arts & Sciences Total		**146**
Business	18-MAY-20	5
Business	19-JAN-20	1
Business	22-MAY-23	27
Business	23-MAY-22	8
Business	24-MAY-21	14
Business Total		**55**
P.C.Rossin Engrg & Applied Sci	18-MAY-20	11
P.C.Rossin Engrg & Applied Sci	19-JAN-20	1
P.C.Rossin Engrg & Applied Sci	22-MAY-23	35
P.C.Rossin Engrg & Applied Sci	23-MAY-22	19
P.C.Rossin Engrg & Applied Sci	24-MAY-21	13
P.C.Rossin Engrg & Applied Sci Total		**79**
Intercollegiate Programs	18-MAY-20	7
Intercollegiate Programs	22-MAY-23	7

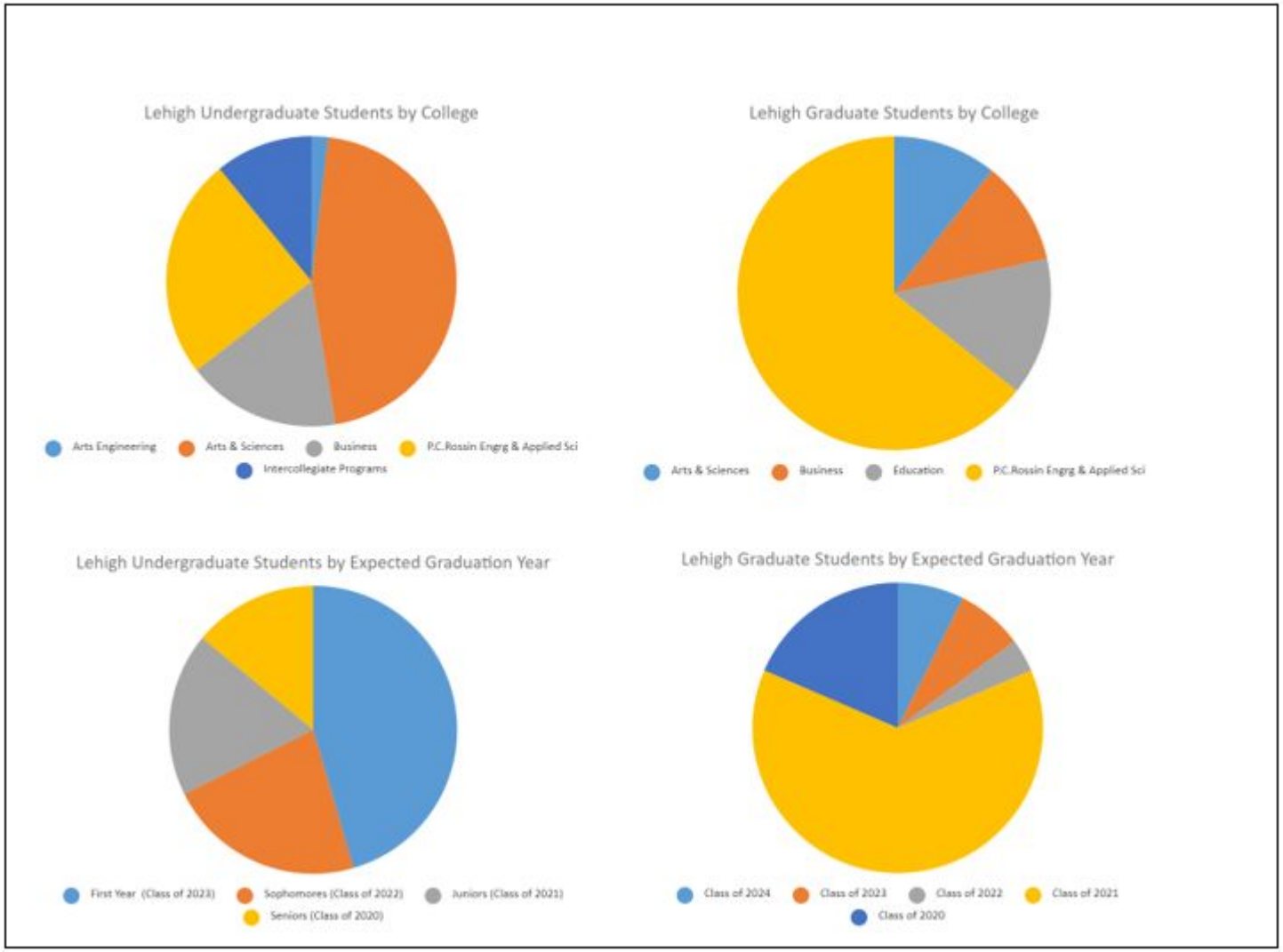
Graduate Level		
College	**Exp Graduation Date**	**Count**
Arts & Sciences	23-MAY-22	1
Arts & Sciences	18-MAY-20	1
Arts & Sciences	24-MAY-21	1
Arts & Sciences Total		**3**
Business	20-MAY-24	1
Business	18-MAY-20	1
Business	(blank)	1
Business Total		**3**
Education	18-MAY-20	2
Education	24-MAY-21	2
Education Total		**4**
P.C.Rossin Engrg & Applied Sci	20-MAY-24	1
P.C.Rossin Engrg & Applied Sci	18-MAY-20	1
P.C.Rossin Engrg & Applied Sci	22-MAY-23	2
P.C.Rossin Engrg & Applied Sci	24-MAY-21	14
P.C.Rossin Engrg & Applied Sci Total		**18**
Grand Total		**28**

Employees Aggregate for Q1

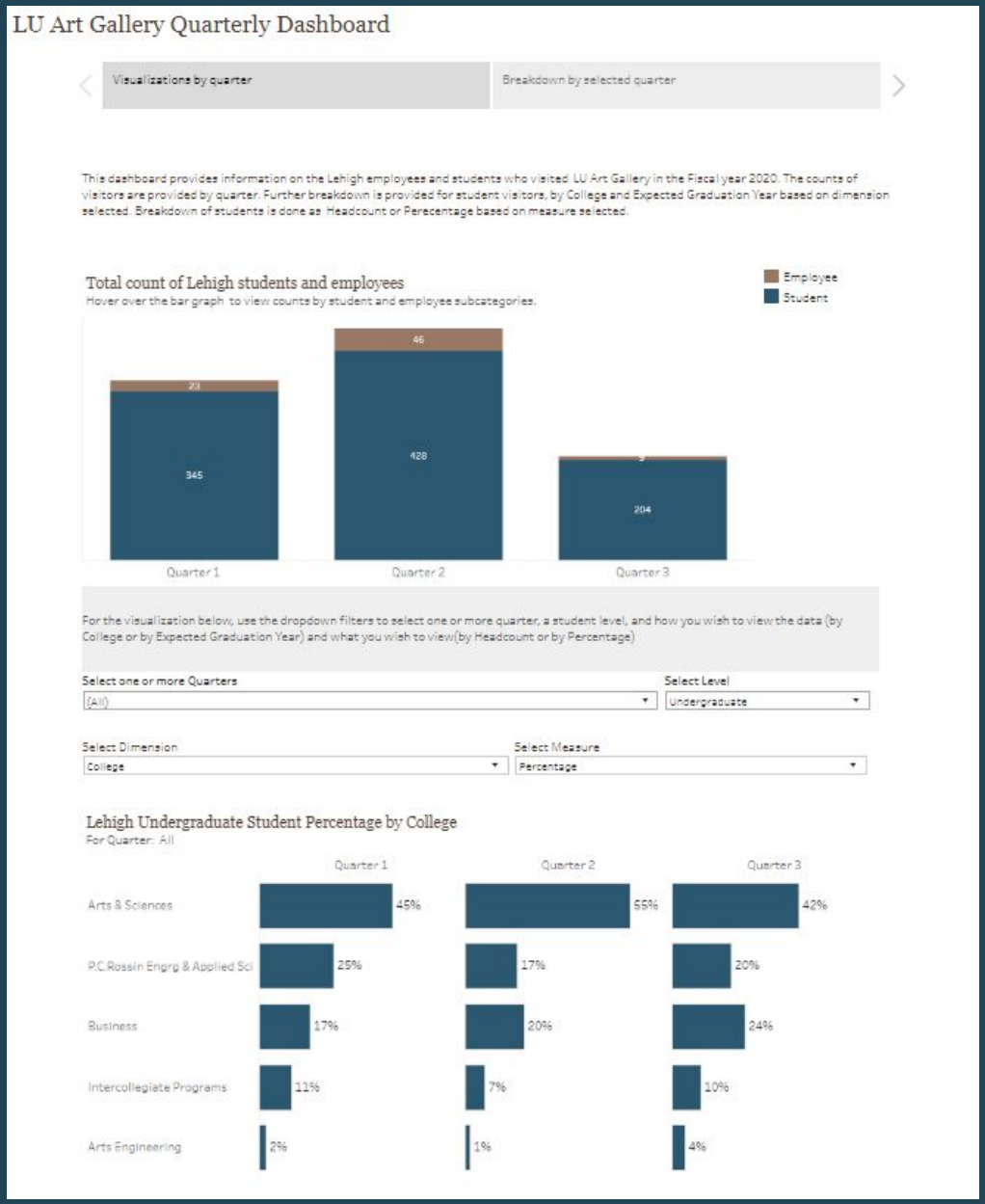
Staff-Exempt and Non exempt	
Home Department	**Count**
Art Galleries	1
Baker Institute	1
BUS UnderGrad Program	1
Communications & Public Affairs	2
Dean Of Students	3
Development and Alumni Relations	5
Information Management	1
Inst Cyber-Physcl Infrastrctr Enrgy	1
Research	1
The Pride Center	1
Grand Total	**17**

Faculty	
Home Department	**Count**
Art, Architecture & Design	1
Computer Science & Engineering	1
Earth & Environmental Sciences	1
English	1
Sociology & Anthropology	1
Grand Total	**5**

LEHIGH UNIVERSITY ART GALLERIES: Q1.FY20



3. Lehigh University Art Gallery - Tableau report



OIRSA | LEHIGH UNIVERSITY
Office of Institutional Research & Strategic Analytics

LU Art Gallery Quarterly Dashboard

Visualizations by quarter | Breakdown by selected quarter

This dashboard provides information on the Lehigh employees and students who visited LU Art Gallery in the Fiscal year 2020. Headcounts are broken down for Lehigh students, faculty & staff for the quarter selected. For students data can be viewed by percentage also.

Select Quarter: Quarter 3

For students, the data is broken down at college level and can be expanded further to expected graduation year.(click + sign next to college name). Select "Undergraduate" or "Graduate" from the level dropdown. Select "Headcount" or "Percentage" from the measure dropdown.

Select Measure: Headcount

Select Level: Undergraduate

Lehigh Undergraduate Student Headcount by College

Quarter 3

Arts & Sciences	Class of 2020	12
	Class of 2021	18
	Class of 2022	23
	Class of 2023	27
Business	Class of 2020	10
	Class of 2021	6
	Class of 2022	20
	Class of 2023	10
P.C.Rossin Engrg & Applied Sci	Class of 2020	9
	Class of 2021	10
	Class of 2022	9
	Class of 2023	9
Intercollegiate Programs	Class of 2020	4
	Class of 2021	2
	Class of 2022	10
	Class of 2023	2

For employees, the data is broken down at home department level. Select "Staff","Faculty" or "Visitng Employee" from the employee type dropdown.

Select Employee Type: Faculty

Lehigh Faculty by Home Department

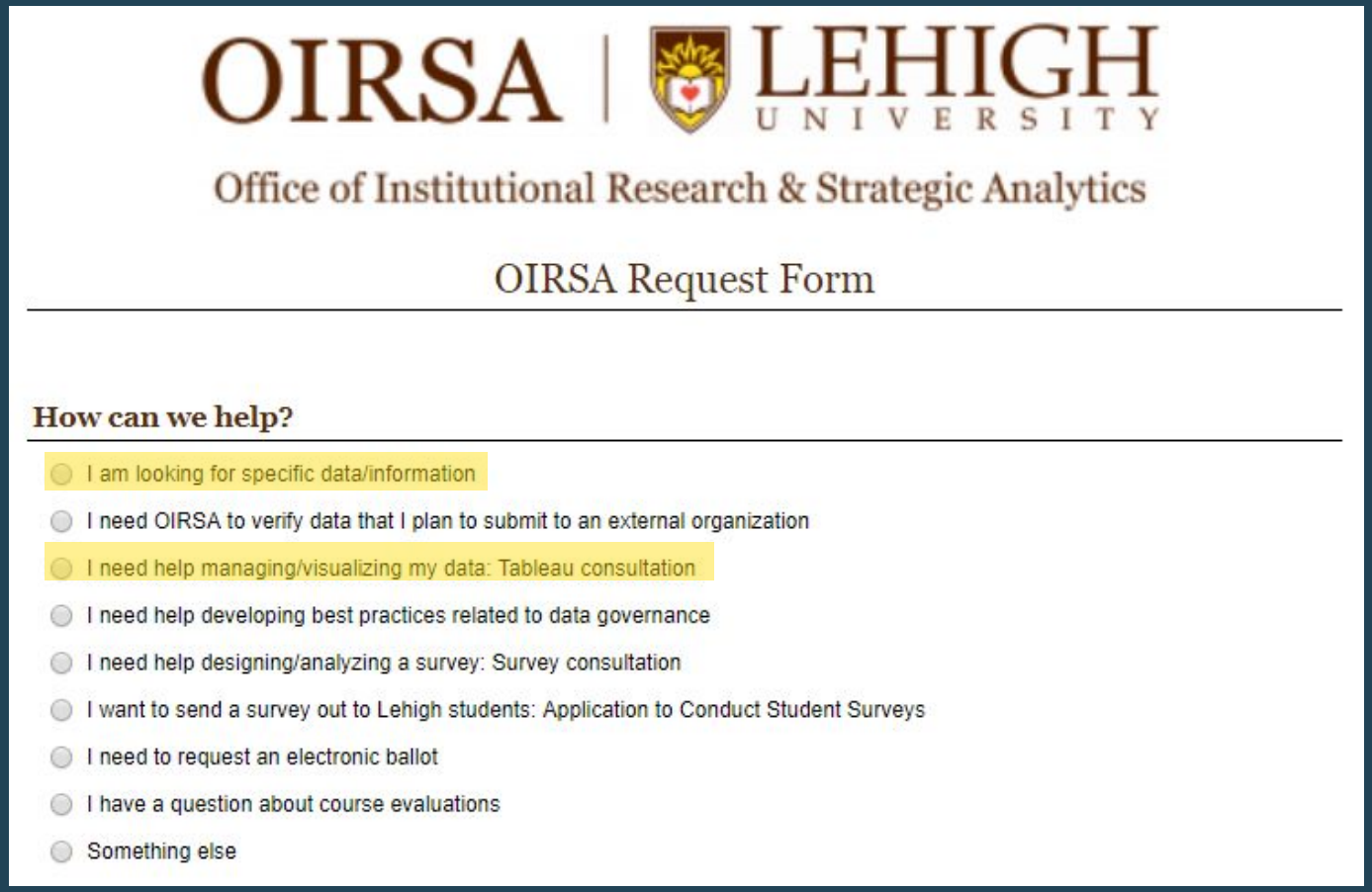
Quarter 3

Art, Architecture & Design	3
Computer Science & Engineering	1
Grand Total	4

OIRSA | LEHIGH UNIVERSITY

Office of Institutional Research & Strategic Analytics

Making a request from OIRSA



OIRSA | LEHIGH UNIVERSITY

Office of Institutional Research & Strategic Analytics

OIRSA Request Form

How can we help?

- I am looking for specific data/information
- I need OIRSA to verify data that I plan to submit to an external organization
- I need help managing/visualizing my data: Tableau consultation
- I need help developing best practices related to data governance
- I need help designing/analyzing a survey: Survey consultation
- I want to send a survey out to Lehigh students: Application to Conduct Student Surveys
- I need to request an electronic ballot
- I have a question about course evaluations
- Something else

Complete the request form at: https://oirsa.lehigh.edu/make-request

OIRSA | LEHIGH UNIVERSITY

Office of Institutional Research & Strategic Analytics

27 Memorial Dr. W. 308 Alumni Memorial Building, Bethlehem PA 18015

610-758-5890
OIR@Lehigh.edu

https://oirsa.lehigh.edu/contact-us

Questions?