Recommended APA Citation:
Le, M. (2015). Visual analytics: Exposing the past, understanding the present, and looking to the future. Workshop Presentation. Presentation at the Annual California Association for Institutional Research Conference, San Francisco, CA.

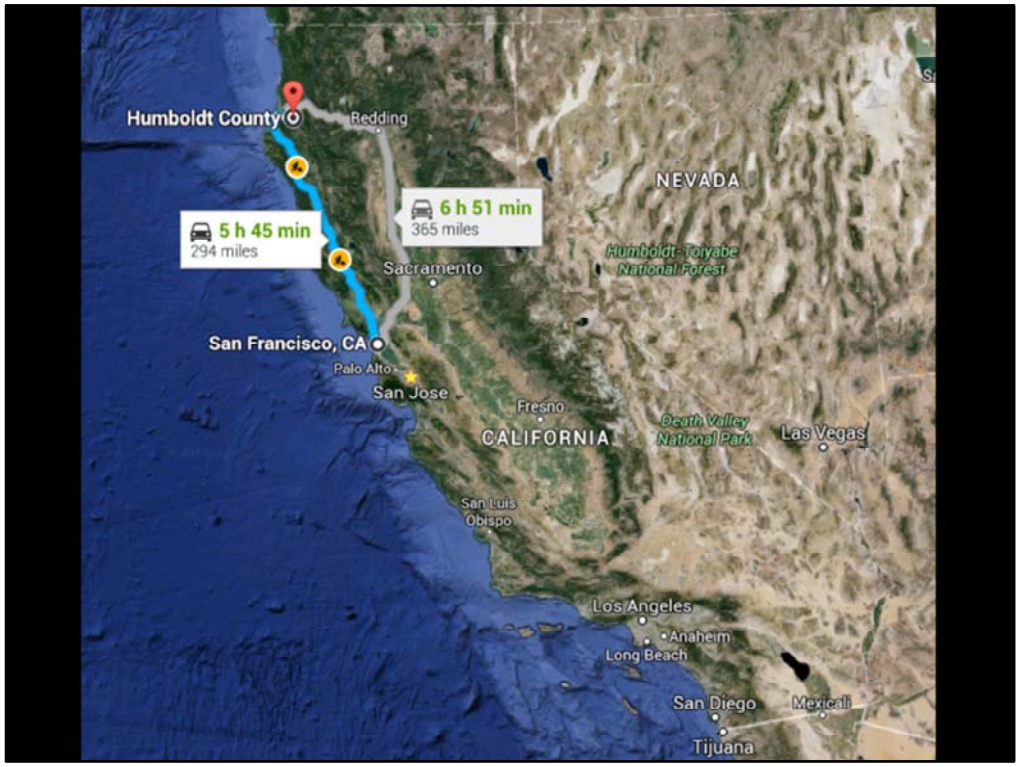
Official Presentation Link:
www.humboldt.edu/irp/presentations.html#2015CAIR

About the Presenter:
https://chroniclevitae.com/people/1091-michael-le

Abstract:
Visual analytics: Exposing the past, understanding the present, and looking to the future
Dan Ariely, founder of The Center for Advanced Hindsight once posted on Facebook, “Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...” This is especially true in Higher Education as much of the work being done to organize, connect, and analyze big data is happening in the for profit sector. This multimedia presentation (video, photos, and text) has three goals. (1) Discuss how the field visual analytics is tackling the problem of analyzing big data. (2) Explore when visual analytics is superior and inferior to typical statistics. (3) Tactics and tools for Institutional Researchers to use in their everyday work to change data into actionable intelligence.

Welcome to Visual Analytics: Exposing the past, understanding the present, and looking to the future. I'm a Research Analyst at Humboldt State University.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



We're about 300 miles North of here, just under the Oregon border.

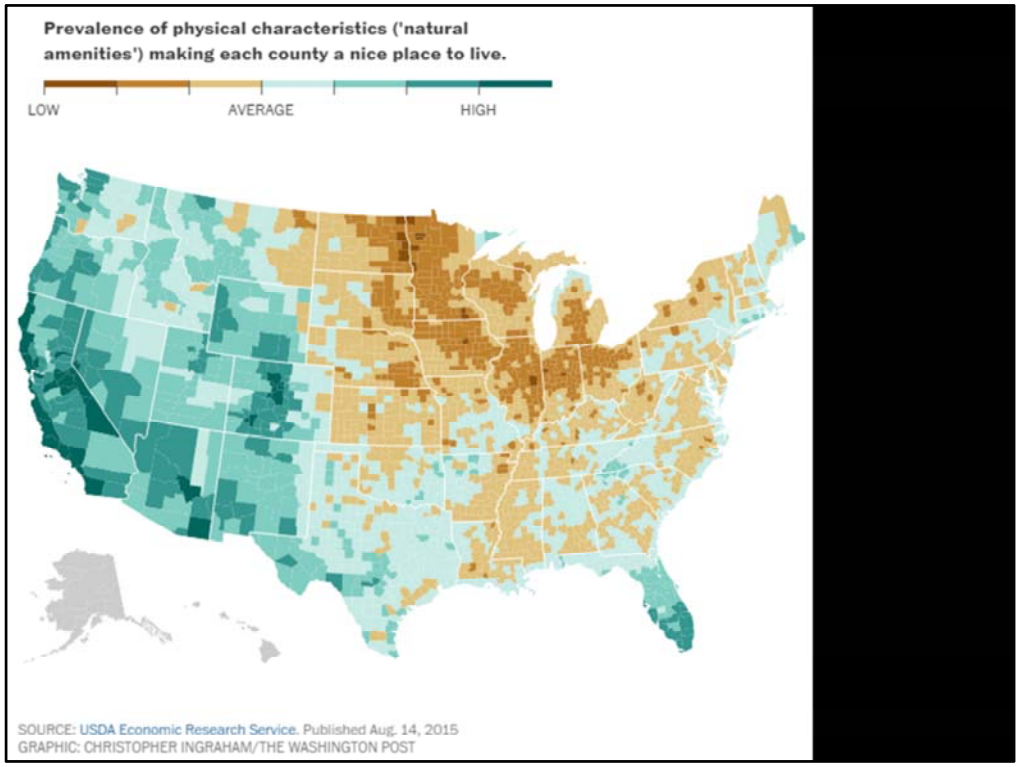
https://www.google.com/maps/dir/San+Francisco,+CA/Humboldt+County,+CA/@36.8913899,-124.3026377,1452819m/am=t/data=!3m1!1e3!4m14!4m13!1m5!1m1!1s0x80859a6d00690021:0x4a501367f076adff!2m2!1d-122.4194155!2d37.7749295!1m5!1m1!1s0x54d3dff475f054b1:0x343a79347db1267f!2m2!1d-123.8695086!2d40.7450055!3e0

Le.(2015). Visual Analytics. CAIR, San Francisco, CA

Who’s heard of Humboldt before?

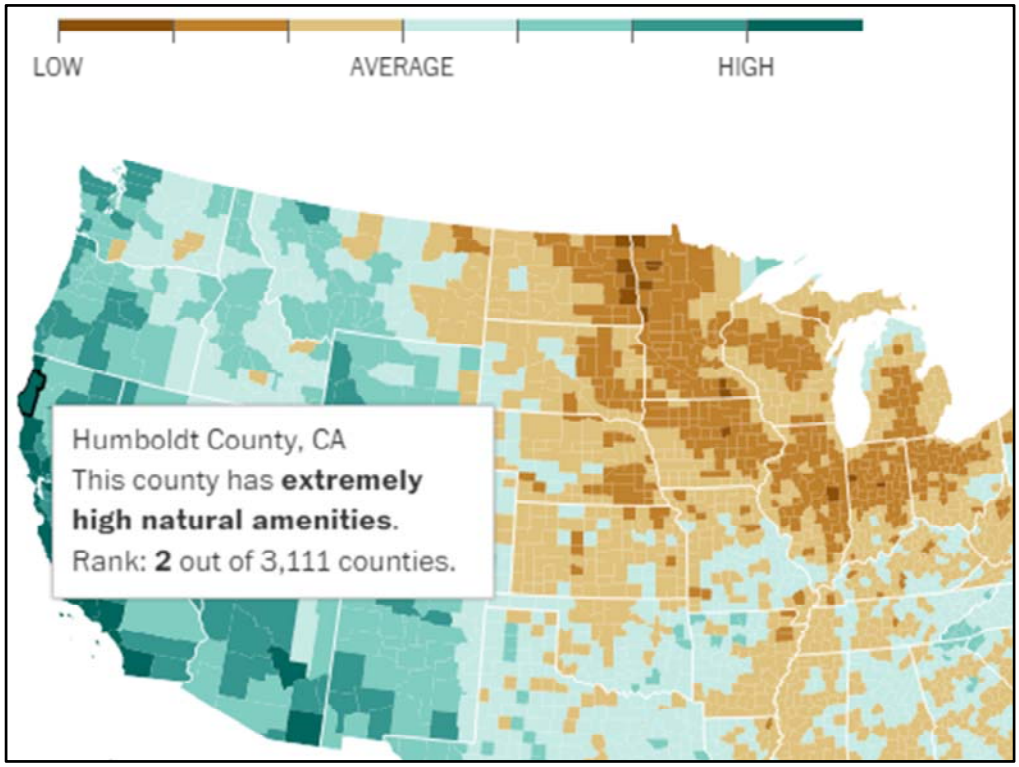
Who's been to Humboldt before? Anyone want to hazard a guess what we're famous for?

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Here's a hint. This is a map of USDA data about physical characteristics of counties in the contiguous United States. Dark green indicates high prevalence of desired natural amenities and dark brown indicates low prevalence.

https://www.washingtonpost.com/news/wonk/wp/2015/08/17/every-county-in-america-ranked-by-natural-beauty/



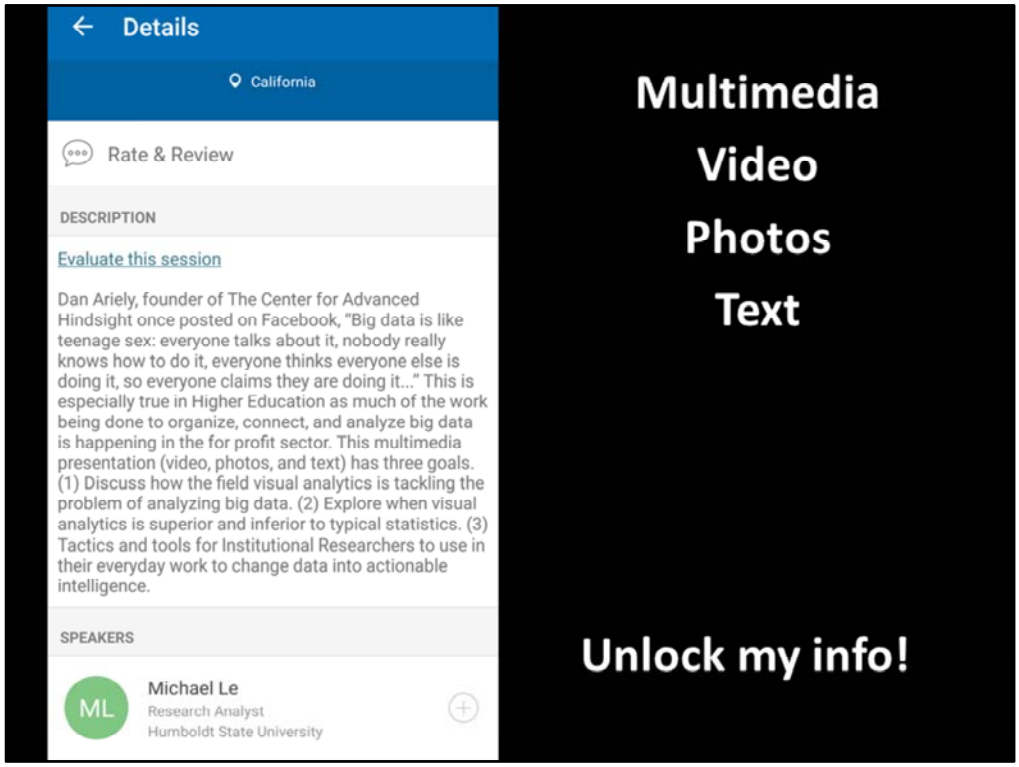
That's right, according to USDA data we're the second most beautiful county in the United States.

The index combines "six measures of climate, topography, and water area that reflect environmental qualities most people prefer." Those qualities, according to the U.S. Department of Agriculture, include mild, sunny winters, temperate summers, low humidity, topographic variation, and access to a body of water.

https://www.washingtonpost.com/news/wonk/wp/2015/08/17/every-county-in-america-ranked-by-natural-beauty/

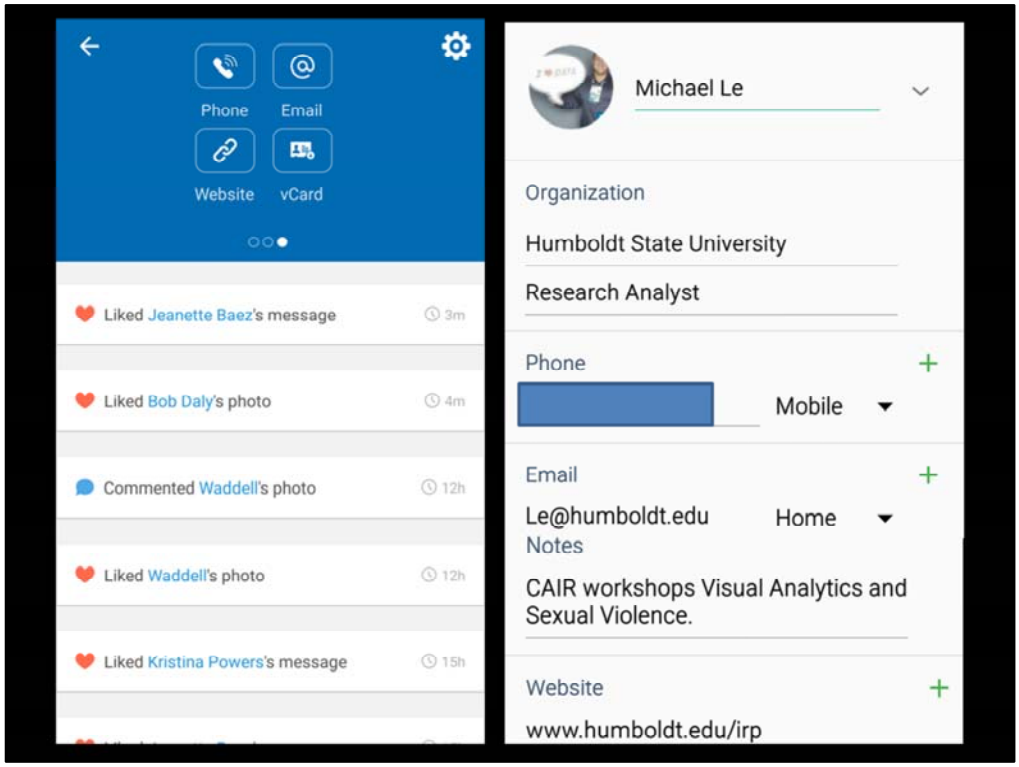
(1) Before we get started, I need to set some house rules.

(2) Raise your hand if you have a tablet, smart phone, or other fancy device with the CAIR app on it.

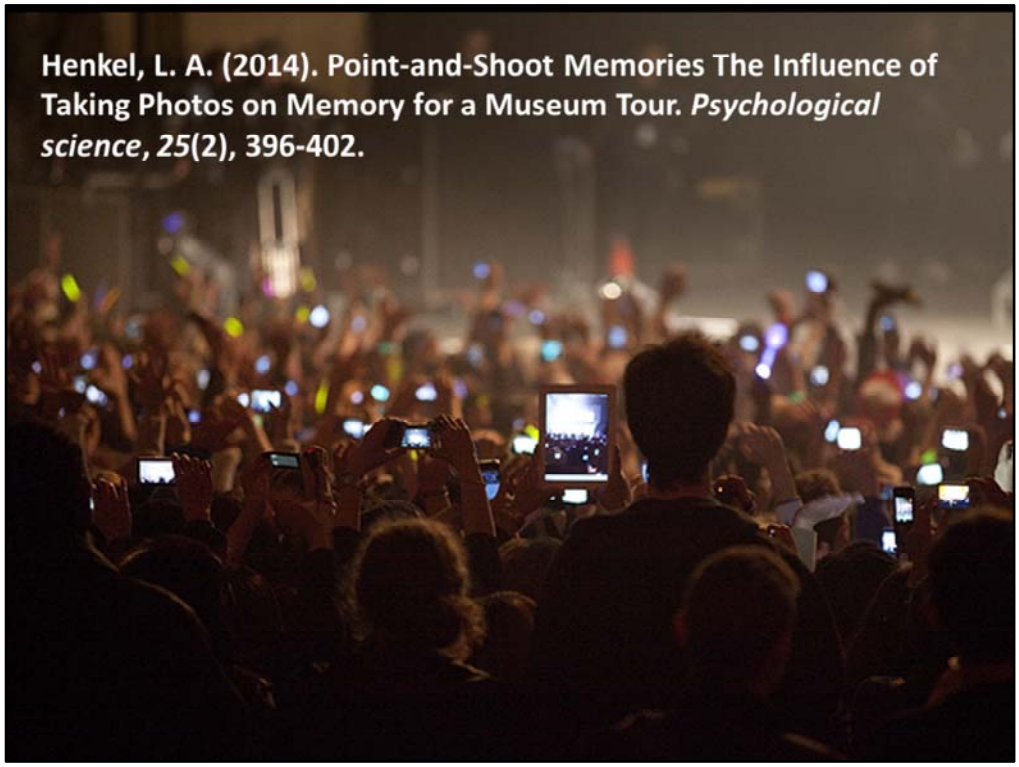


Now keep your hand up if you plan to complete an evaluation on your fancy device. Everyone else, please make sure to complete a paper evaluation. I would appreciate your comments on, is what you thought of my use videos, photos, and text. The only way I get better as a presenter is if you give me honest critique and praise. As a door prize for completing the evaluation, you can unlock my Full contact information in under the speakers portion of the app.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Just Joking, just click on my name, go to vcard, and you can add it to your Google Contacts. Apple people – I have no idea how this works for you.



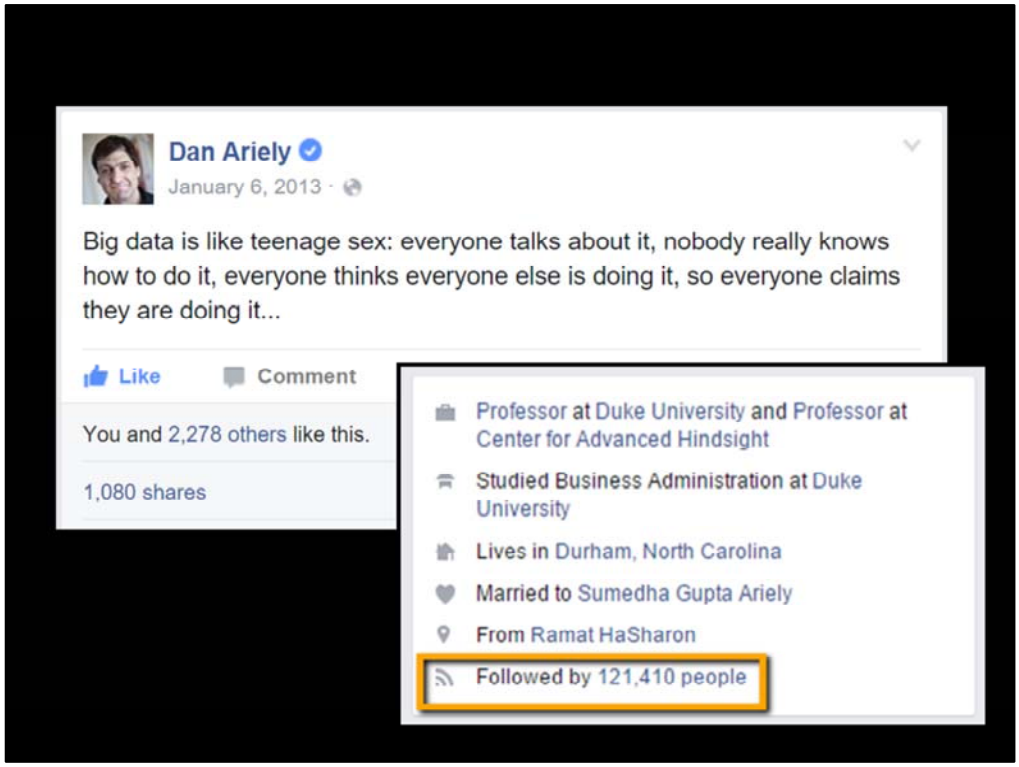
Several studies have shown that by taking a photo you offload the task of memorizing something to your device. You are less likely to recall the content at a later time. Henkel, (2014) setup a study where they led participants on a guided tour of an art museum and directed participants to observe some objects and to photograph others. Results showed a photo-taking-impairment effect: If participants took a photo of each object as a whole, they remembered fewer objects and remembered fewer details about the objects and the objects' locations in the museum than if they instead only observed the objects and did not photograph them.

Henkel, L. A. (2014). Point-and-Shoot Memories The Influence of Taking Photos on Memory for a Museum Tour. *Psychological science*, *25*(2), 396-402.

But if you must take a photo, this is the one slide worth photo

Download my full PowerPoint presentation with my full-text which is likely to be more detailed and inclusive than the actual presentation.

In the next 40 minutes, I have three goals: (1) Discuss how the field visual analytics is tackling the problem of analyzing big data. (2) Explore when visual analytics is superior and inferior to typical statistics. (3) Tactics and tools for Institutional Researchers to use in their everyday work to change data into actionable intelligence.



Dan Ariely, founder of The Center for Advanced Hindsight posted on Facebook in 2013, “Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...”

https://www.facebook.com/dan.ariely

While we're talking about Facebook, do you know Facebook uses blue as it's main color? For Mark Zuckerberg, blue is the "richest" color he can see — which is why Facebook's dominant color is blue.

http://factually.gizmodo.com/facebook-is-blue-because-mark-zuckerberg-is-colorblind-1598739229

Red-green color vision defects are the most common form of color vision deficiency. This condition affects males much more often than females. Among populations with Northern European ancestry, it occurs in about 1 in 12 males and 1 in 200 females. Blue-yellow color vision defects affect males and females equally. This condition occurs in fewer than 1 in 10,000 people worldwide.

http://ghr.nlm.nih.gov/condition/color-vision-deficiency

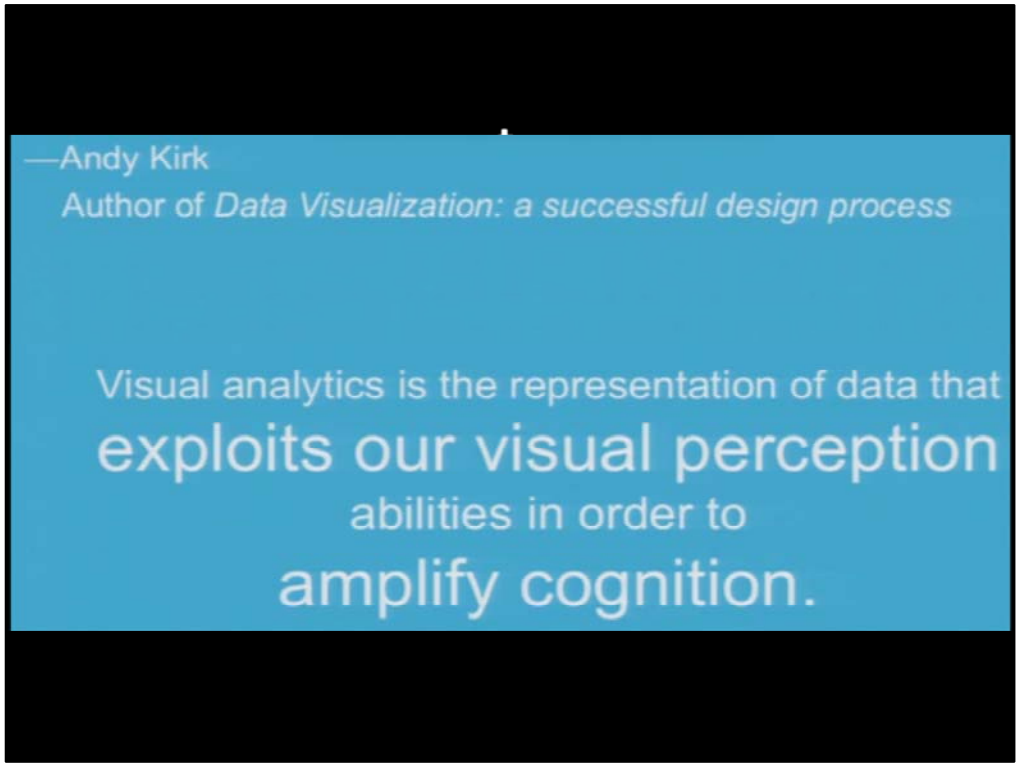
I digress, back to my buddy Dan. He's not really my buddy – I just follow him on Facebook. I think this quote is especially true in Higher Education. Much of the work being done to organize, connect, and analyze big data is happening in the for profit sector.

http://www.ted.com/speakers/dan_ariely

What is big data? Well, watch this video and find out.

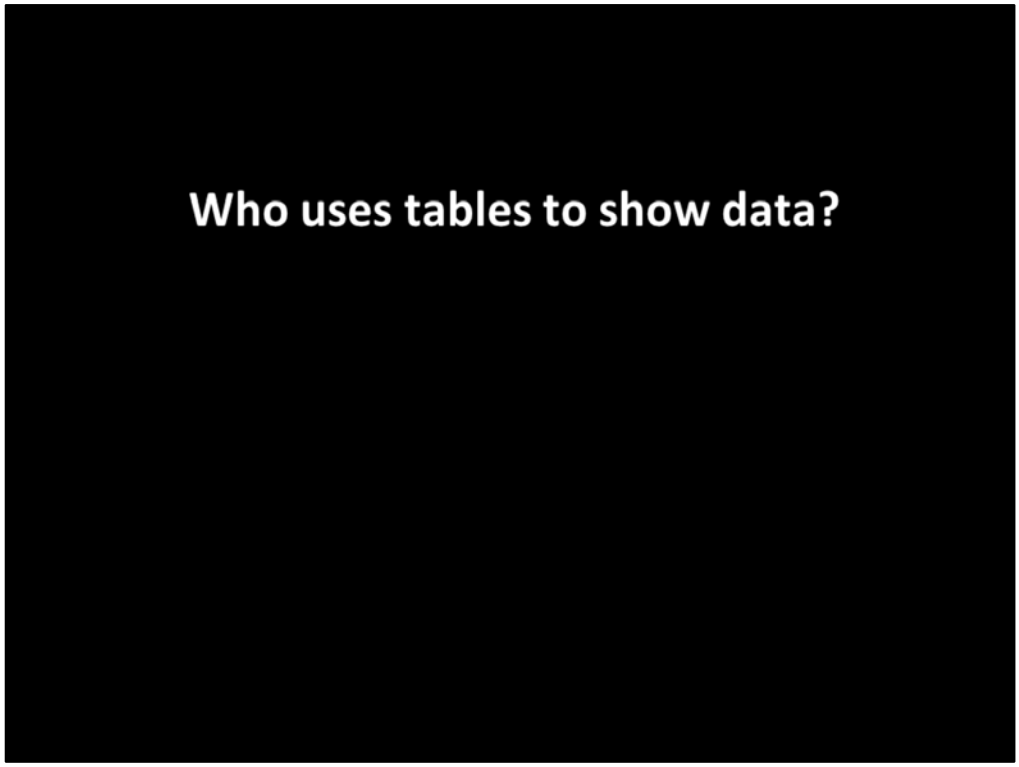
https://www.youtube.com/watch?v=TzxmjbL-i4Y

Is anyone here using big data? I certainly am not, but I sure would like to. I am however using tactics that analyst are using in the “for profit” world to analyze data.



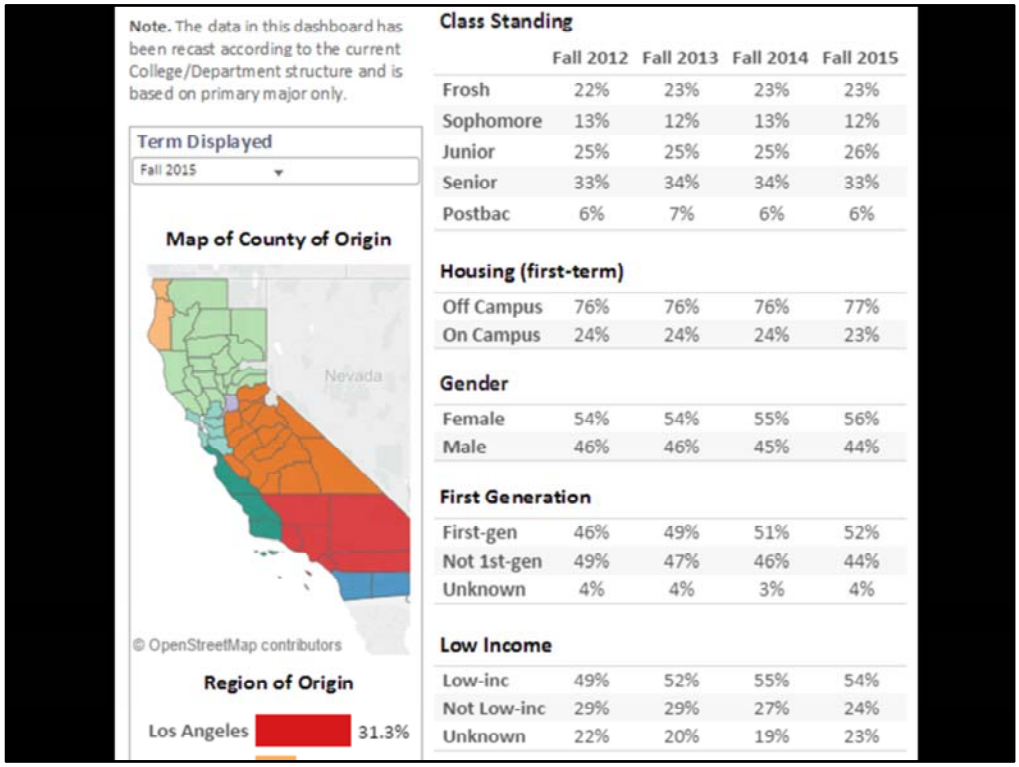
The goal of visual analytics is to exploit the visual system and amplify cognition. Cognition being the mental processing of a chart, table, or other representation of data.

http://www.amazon.com/Andy-Kirk/e/B00J39EBMW/ref=dp_byline_cont_book_1



Who uses tables to show data? I know I do.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



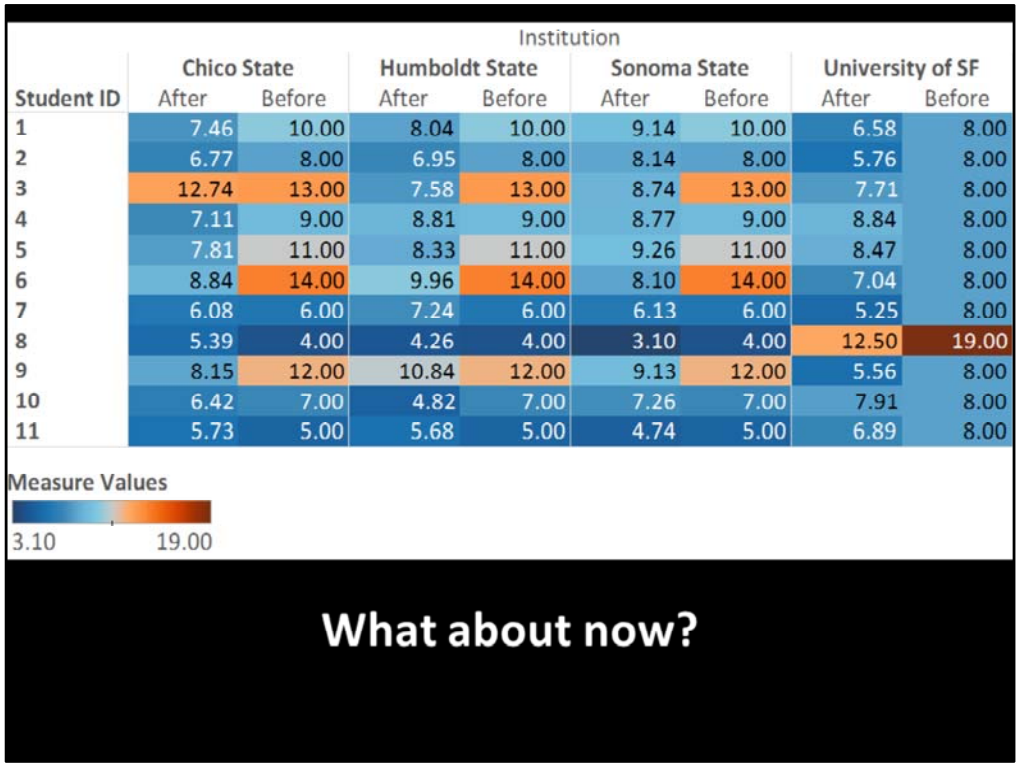
Here is one of the first dashboards I built. Three editions later and still mostly has just tables. Tables can be very useful for presenting data.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA

	Institution							
	Chico State		Humboldt State		Sonoma State		University of SF	
Student ID	After	Before	After	Before	After	Before	After	Before
1	7.46	10.00	8.04	10.00	9.14	10.00	6.58	8.00
2	6.77	8.00	6.95	8.00	8.14	8.00	5.76	8.00
3	12.74	13.00	7.58	13.00	8.74	13.00	7.71	8.00
4	7.11	9.00	8.81	9.00	8.77	9.00	8.84	8.00
5	7.81	11.00	8.33	11.00	9.26	11.00	8.47	8.00
6	8.84	14.00	9.96	14.00	8.10	14.00	7.04	8.00
7	6.08	6.00	7.24	6.00	6.13	6.00	5.25	8.00
8	5.39	4.00	4.26	4.00	3.10	4.00	12.50	19.00
9	8.15	12.00	10.84	12.00	9.13	12.00	5.56	8.00
10	6.42	7.00	4.82	7.00	7.26	7.00	7.91	8.00
11	5.73	5.00	5.68	5.00	4.74	5.00	6.89	8.00

But then again, sometimes not. What stands out to about this data table? This is fictitious data about Chico State, Humboldt State, Sonoma State, and the University of San Francisco. We're comparing before and after math test scores of 10 students at each university.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



	Institution							
	Chico State		Humboldt State		Sonoma State		University of SF	
Student ID	After	Before	After	Before	After	Before	After	Before
1	7.46	10.00	8.04	10.00	9.14	10.00	6.58	8.00
2	6.77	8.00	6.95	8.00	8.14	8.00	5.76	8.00
3	12.74	13.00	7.58	13.00	8.74	13.00	7.71	8.00
4	7.11	9.00	8.81	9.00	8.77	9.00	8.84	8.00
5	7.81	11.00	8.33	11.00	9.26	11.00	8.47	8.00
6	8.84	14.00	9.96	14.00	8.10	14.00	7.04	8.00
7	6.08	6.00	7.24	6.00	6.13	6.00	5.25	8.00
8	5.39	4.00	4.26	4.00	3.10	4.00	12.50	19.00
9	8.15	12.00	10.84	12.00	9.13	12.00	5.56	8.00
10	6.42	7.00	4.82	7.00	7.26	7.00	7.91	8.00
11	5.73	5.00	5.68	5.00	4.74	5.00	6.89	8.00

Simply coding the table with color helps us process the data better. Now you should be able to easily pick out the outliers and find you min and max scores.

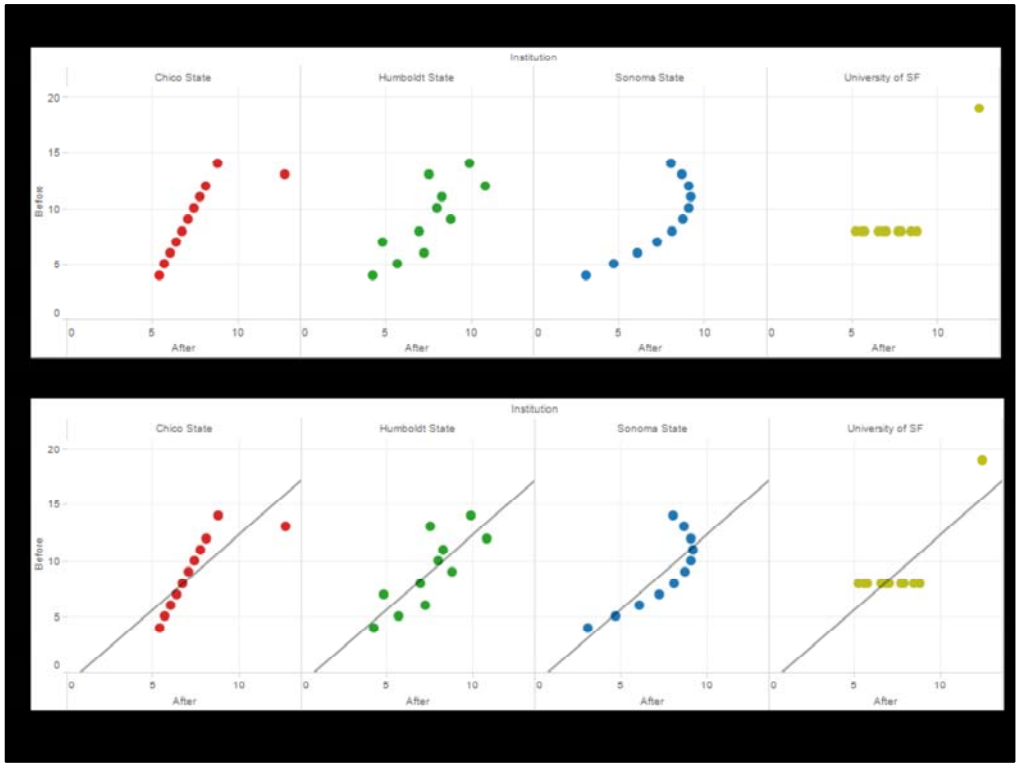
	Institution							
	Chico State		Humboldt State		Sonoma State		University of SF	
Student ID	After	Before	After	Before	After	Before	After	Before
1	7.46	10.00	8.04	10.00	9.14	10.00	6.58	8.00
2	6.77	8.00	6.95	8.00	8.14	8.00	5.76	8.00
3	12.74	13.00	7.58	13.00	8.74	13.00	7.71	8.00
4	7.11	9.00	8.81	9.00	8.77	9.00	8.84	8.00
5	7.81	11.00	8.33	11.00	9.26	11.00	8.47	8.00
6	8.84	14.00	9.96	14.00	8.10	14.00	7.04	8.00
7	6.08	6.00	7.24	6.00	6.13	6.00	5.25	8.00
8	5.39	4.00	4.26	4.00	3.10	4.00	12.50	19.00
9	8.15	12.00	10.84	12.00	9.13	12.00	5.56	8.00
10	6.42	7.00	4.82	7.00	7.26	7.00	7.91	8.00
11	5.73	5.00	5.68	5.00	4.74	5.00	6.89	8.00

	Institution			
	Chico State	Humboldt State	Sonoma State	University of SF
Avg. Before	9.000	9.000	9.000	9.000
Variance of Before	11.000	11.000	11.000	11.000
Avg. After	7.500	7.501	7.501	7.501
Variance of After	4.123	4.127	4.128	4.123

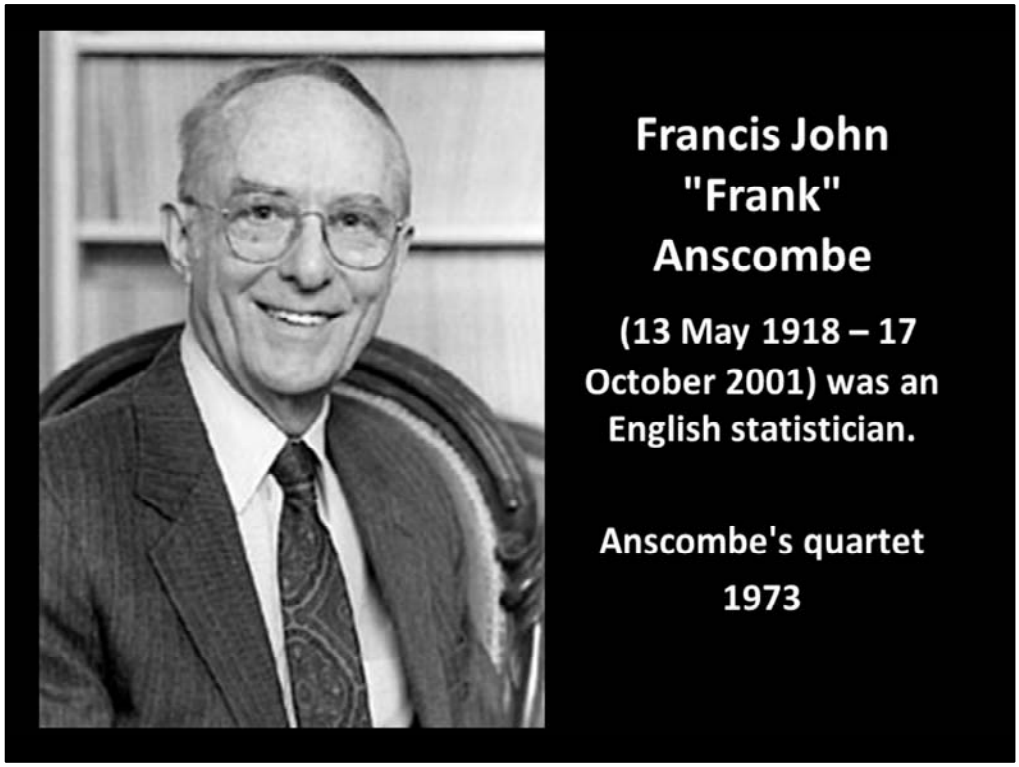
Let's not forget about our statistics. When you have a before and after measure of scale variables (i.e.., the data is interval/ratio, not categorical), we might want to run a Regression Analysis. Before doing that, we should examine our mean and variance to make sure we our data will be good to model.

Humm… that's interesting, the before and after mean and variance are the same across the 4 groups.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA

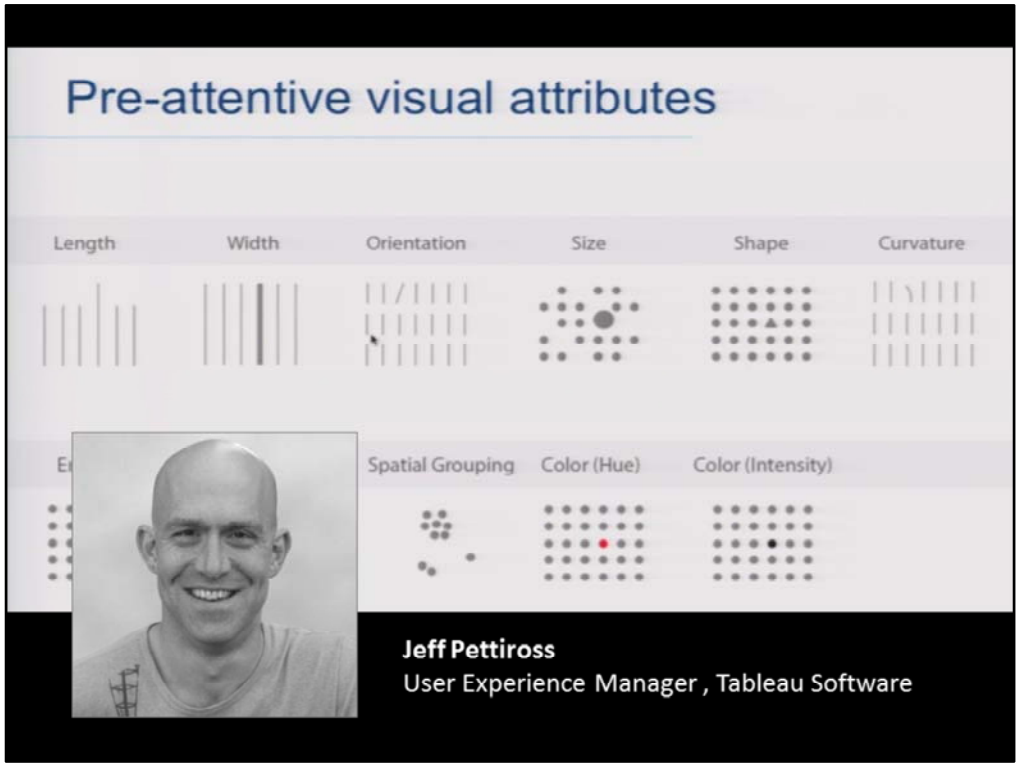


If we visualize the data, we can see interesting things happening in the data that we would have missed had we not visualized the data in several different ways. We definitely did not see these patterns in the table view.



If you're wondering about this data set, it comes from Frank Anscombe and is known as the Anscombe's quartet of 1973.

http://data.heapanalytics.com/anscombes-quartet-and-why-summary-statistics-dont-tell-the-whole-story/
https://en.wikipedia.org/wiki/Anscombe%27s_quartet
https://en.wikipedia.org/wiki/Frank_Anscombe
http://www.statsblogs.com/2015/02/26/using-and-abusing-data-visualization-anscombes-quartet-and-cheating-bonferroni/
http://data.heapanalytics.com/anscombes-quartet-and-why-summary-statistics-dont-tell-the-whole-story/

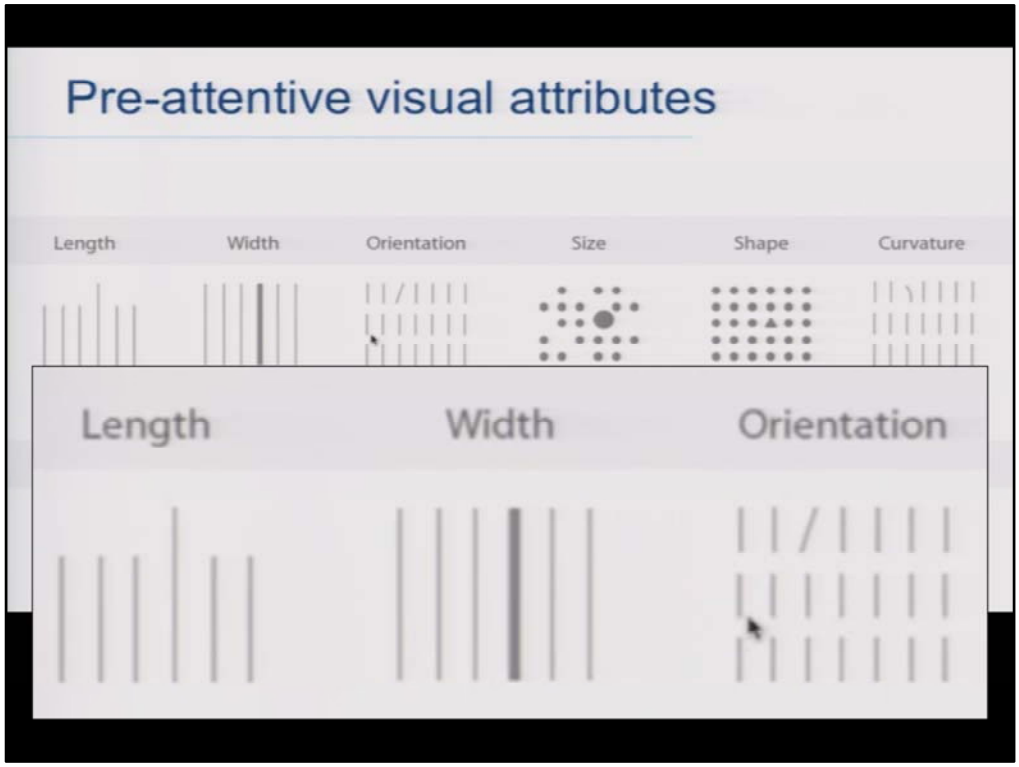


Jeff has been dreaming about information design and user experience since he was a kid, spending hours sketching out interface elements for games he would write on his TRS-80. Since then, he has spent the last 25 years designing software and information visualizations in diverse fields including, data analysis and visualization, healthcare, project management, desktop software, and engineering. As User Experience Manager, he races to keep up with the great ideas about flow generated by the rest of Tableau.

EASTER EGG – Watch Jeff's full presentation for free at:
https://tc14.tableau.com/schedule/content/1172

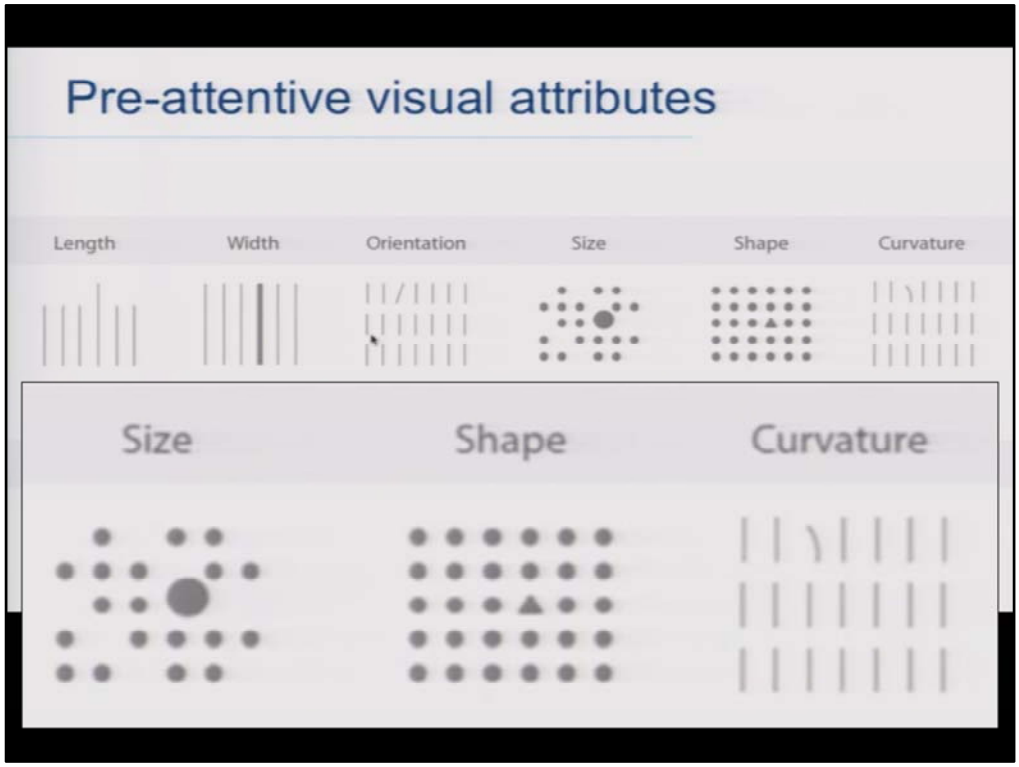
SECOND EASTER EGG – Read Tapping the Power of Visual Perception by Stephen Few
https://www.perceptualedge.com/articles/ie/visual_perception.pdf

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



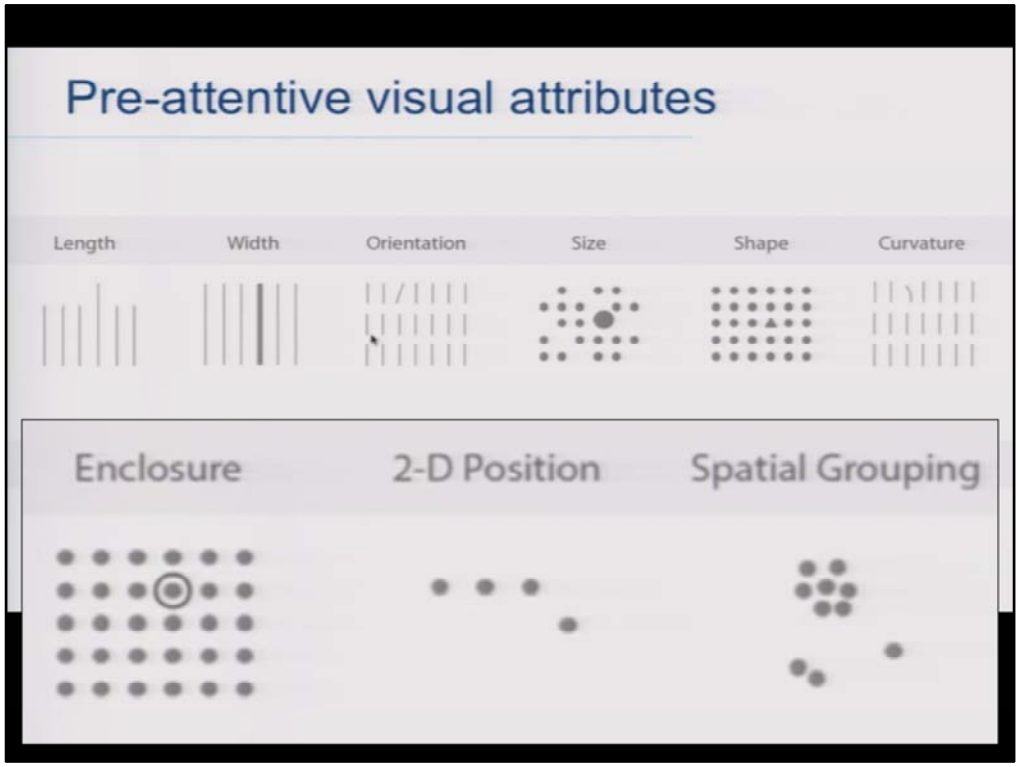
Length, width, and orientation

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



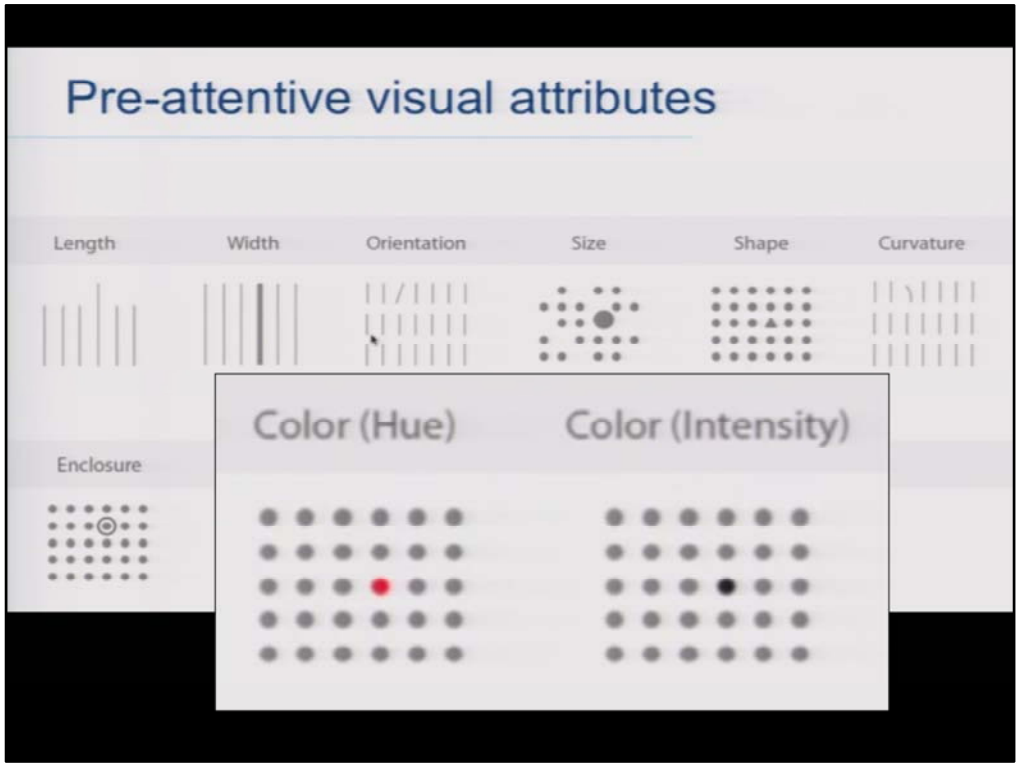
Size, Shape, and Curvature

Le.(2015). Visual Analytics. CAIR, San Francisco, CA

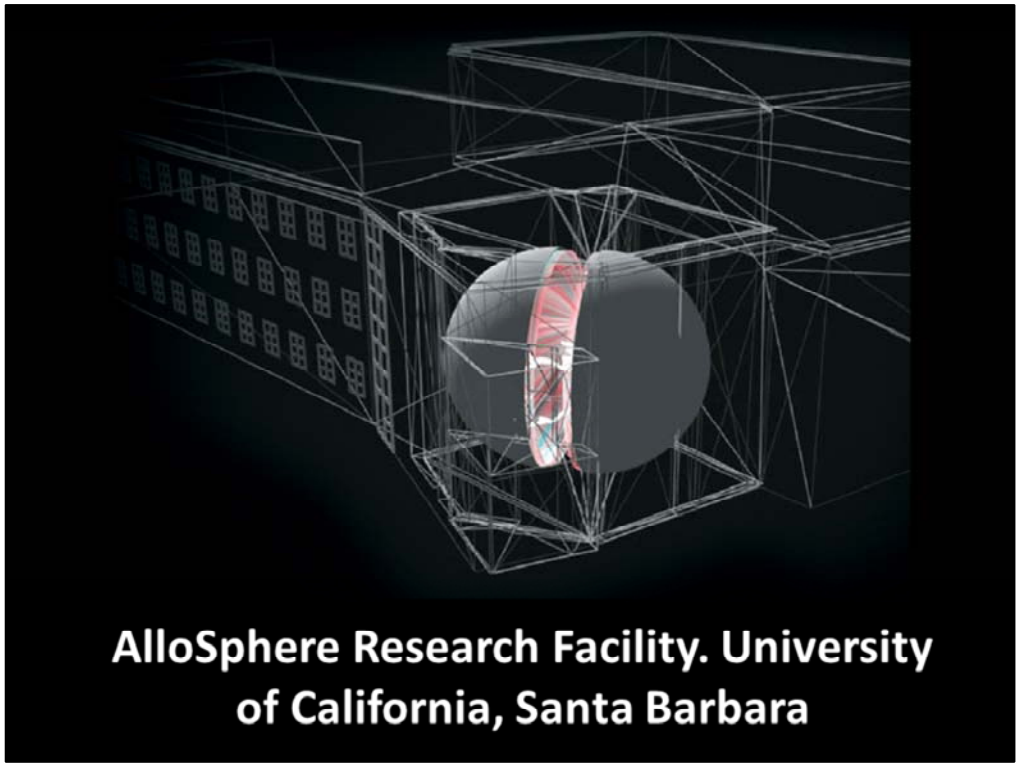


Enclosure – which I think it just shape and size, 2-D Position, and Spatial Grouping.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Two that I use in almost every dashboard are



Who has heard of the AlloSphere? The AlloSphere is a new way to see, hear and interpret scientific data. Dive into the brain, feel electron spin, hear the music of the elements ... and detect previously unseen patterns that could lead to new discoveries.

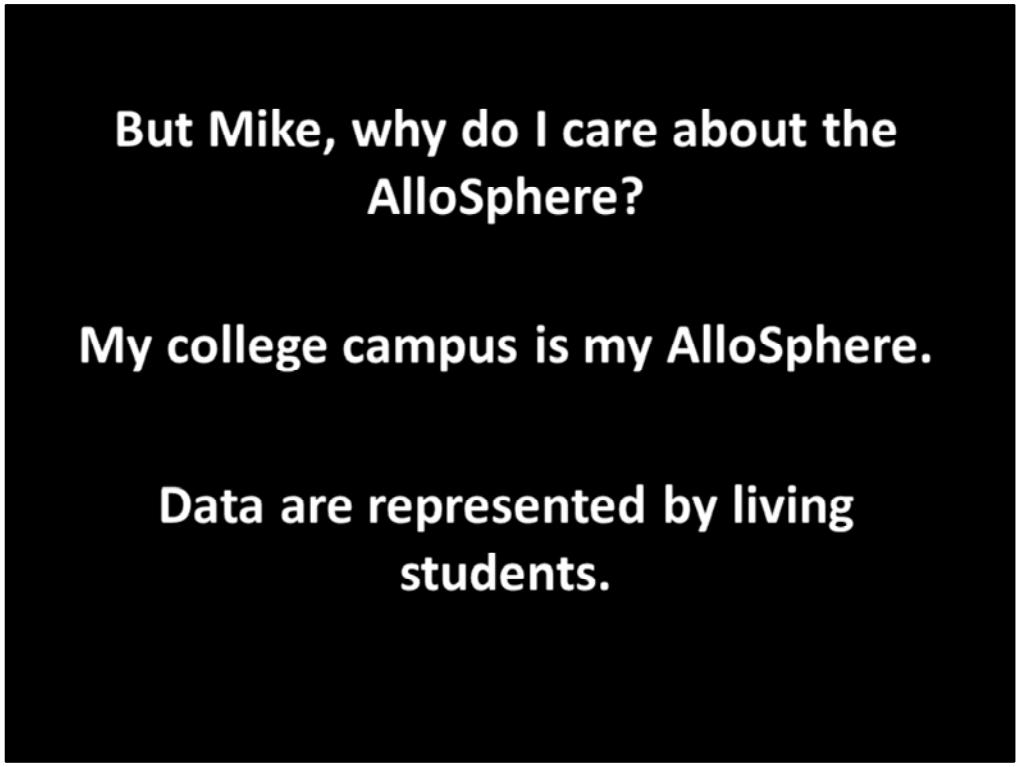
http://www.allosphere.ucsb.edu/index.php

Le.(2015). Visual Analytics. CAIR, San Francisco, CA

Video found at:

https://www.ted.com/talks/joann_kuchera_morin_tours_the_allosphere?language=en

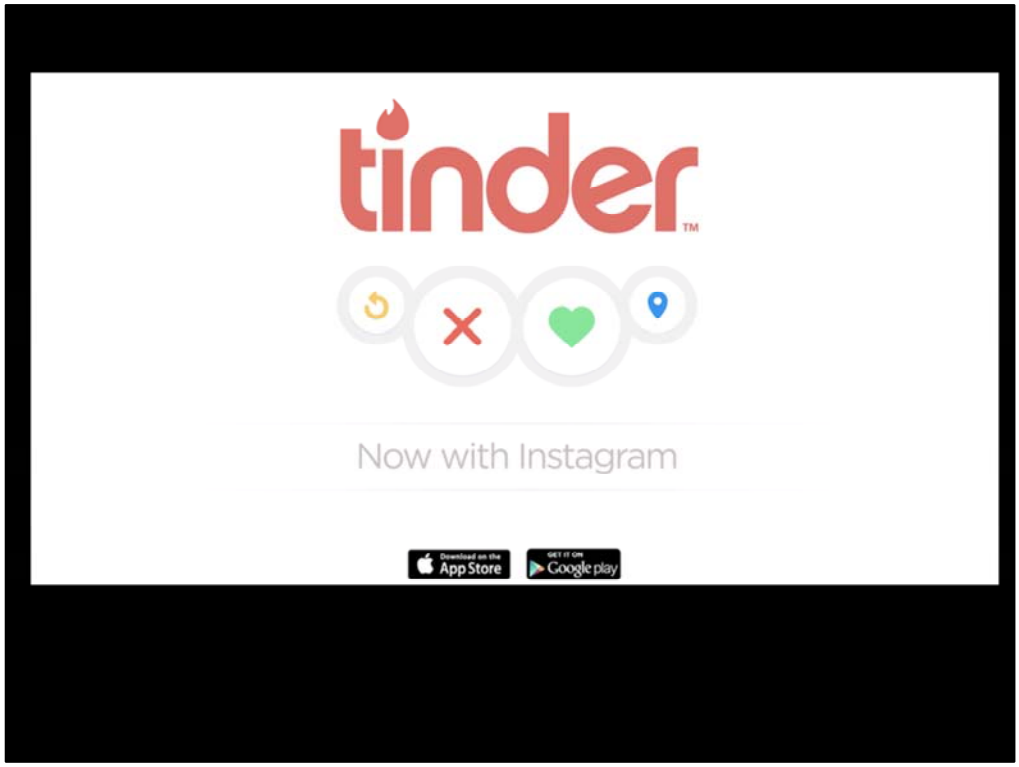
Keep in mind, this video was filmed in 2009… I wonder what they can do now?



But Mike, why do I care about the AlloSphere?

My college campus is my AlloSphere.

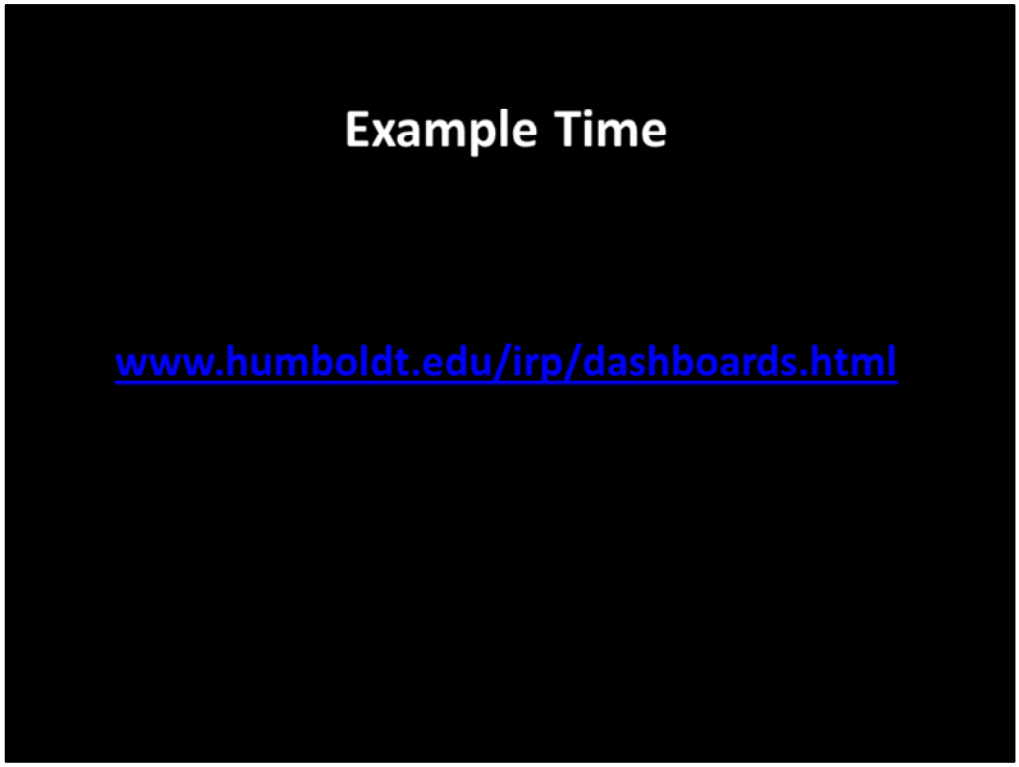
Data are represented by living students.



But not all of us can go into a 3d sphere to visualize our data. So what can we do. Well, just like how tinder brought meeting and dating people to an app where you swipe right to keep and left to pass, there's no a dating app for Data! Let me show you vizable, the next evolution of drag and drop analytics.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA

https://vizable.tableau.com/



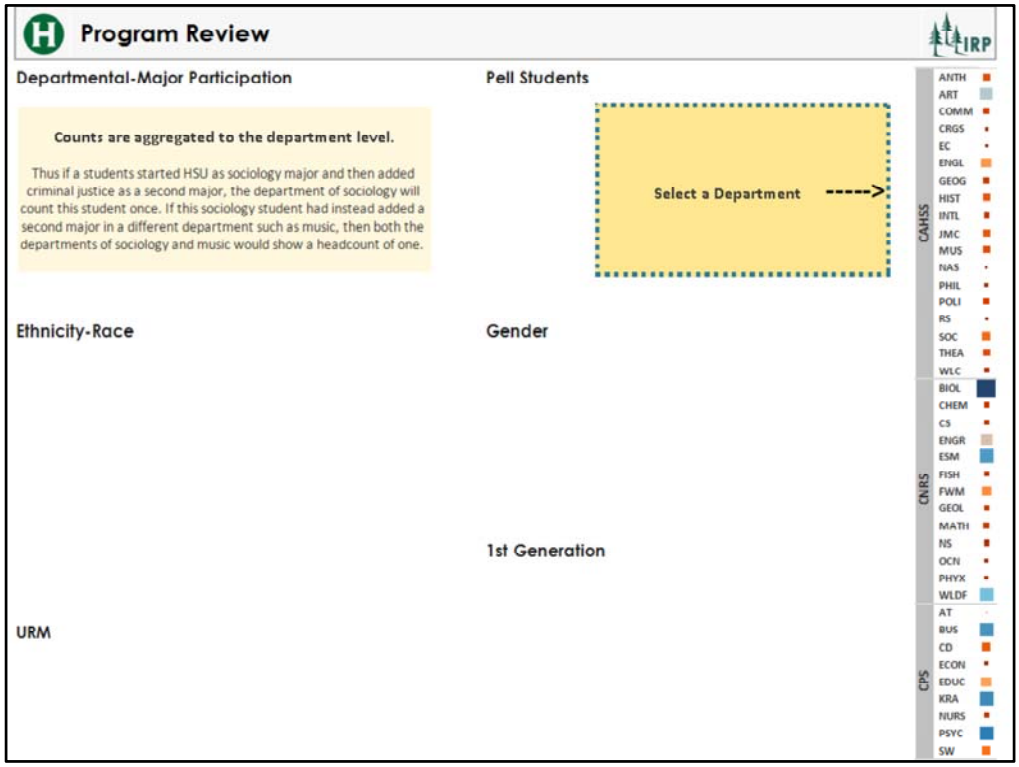
While all public Dashboards can be found at: www.humboldt.edu/irp/dashboards.html, I'm going to show you the beta version of our Program Review Workbook that will be released in January, but not available publically. It was designed using Tableau Desktop (Paid Product), shared via Google Drive (Free), and accessed by users through Tableau Reader (Free).

Le.(2015). Visual Analytics. CAIR, San Francisco, CA

Page 1

Provide short definitions for terms used in the workbook.

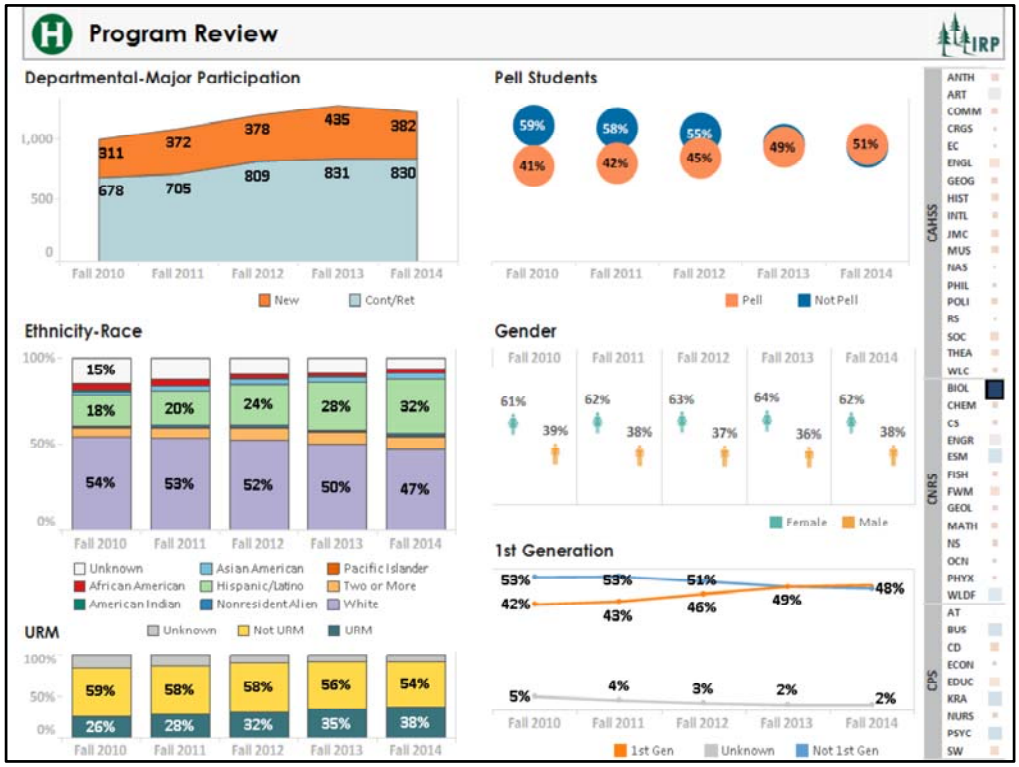
Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Page 2

Departments level metrics that is not based on university headcount. Instead, this shows headcount at the Department level accounting for second and third majors.

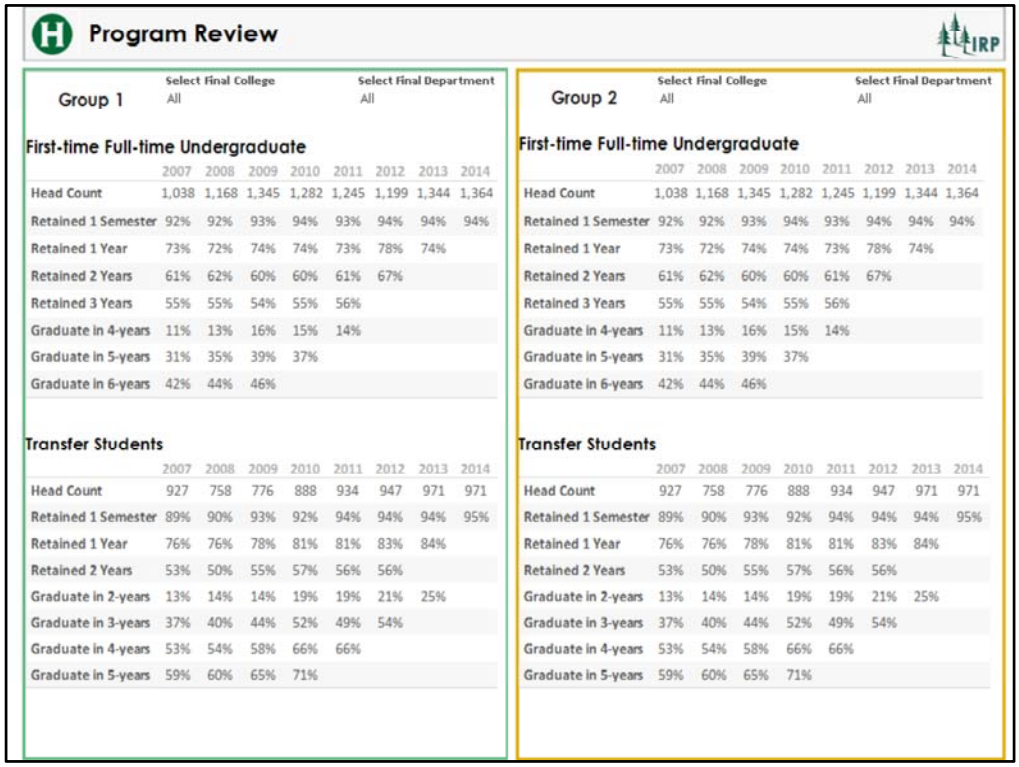
Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Page 2

Biology Department Example.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Program Review

Group 1

Select Final College: All

Select Final Department: All

First-time Full-time Undergraduate

	2007	2008	2009	2010	2011	2012	2013	2014
Head Count	1,038	1,168	1,345	1,282	1,245	1,199	1,344	1,364
Retained 1 Semester	92%	92%	93%	94%	93%	94%	94%	94%
Retained 1 Year	73%	72%	74%	74%	73%	78%	74%	
Retained 2 Years	61%	62%	60%	60%	61%	67%		
Retained 3 Years	55%	55%	54%	55%	56%			
Graduate in 4-years	11%	13%	16%	15%	14%			
Graduate in 5-years	31%	35%	39%	37%				
Graduate in 6-years	42%	44%	46%					

Transfer Students

	2007	2008	2009	2010	2011	2012	2013	2014
Head Count	927	758	776	888	934	947	971	971
Retained 1 Semester	89%	90%	93%	92%	94%	94%	94%	95%
Retained 1 Year	76%	76%	78%	81%	81%	83%	84%	
Retained 2 Years	53%	50%	55%	57%	56%	56%		
Graduate in 2-years	13%	14%	14%	19%	19%	21%	25%	
Graduate in 3-years	37%	40%	44%	52%	49%	54%		
Graduate in 4-years	53%	54%	58%	66%	66%			
Graduate in 5-years	59%	60%	65%	71%				

Group 2

Select Final College: All

Select Final Department: All

First-time Full-time Undergraduate

	2007	2008	2009	2010	2011	2012	2013	2014
Head Count	1,038	1,168	1,345	1,282	1,245	1,199	1,344	1,364
Retained 1 Semester	92%	92%	93%	94%	93%	94%	94%	94%
Retained 1 Year	73%	72%	74%	74%	73%	78%	74%	
Retained 2 Years	61%	62%	60%	60%	61%	67%		
Retained 3 Years	55%	55%	54%	55%	56%			
Graduate in 4-years	11%	13%	16%	15%	14%			
Graduate in 5-years	31%	35%	39%	37%				
Graduate in 6-years	42%	44%	46%					

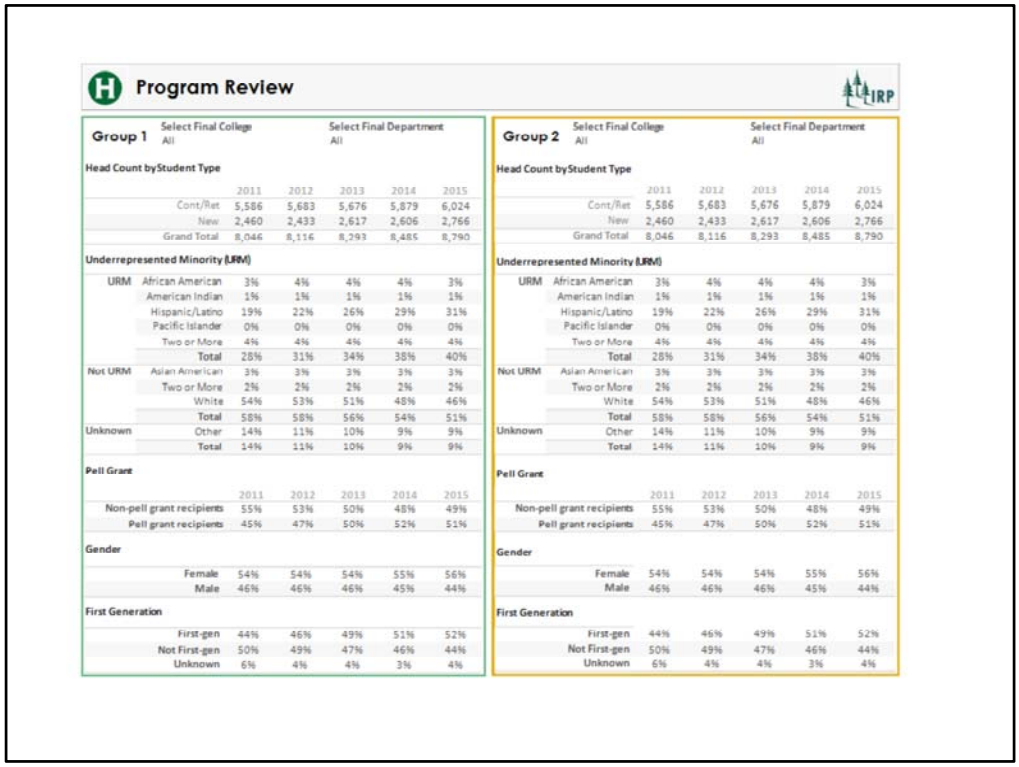
Transfer Students

	2007	2008	2009	2010	2011	2012	2013	2014
Head Count	927	758	776	888	934	947	971	971
Retained 1 Semester	89%	90%	93%	92%	94%	94%	94%	95%
Retained 1 Year	76%	76%	78%	81%	81%	83%	84%	
Retained 2 Years	53%	50%	55%	57%	56%	56%		
Graduate in 2-years	13%	14%	14%	19%	19%	21%	25%	
Graduate in 3-years	37%	40%	44%	52%	49%	54%		
Graduate in 4-years	53%	54%	58%	66%	66%			
Graduate in 5-years	59%	60%	65%	71%				

Page 3

Compare side by side outcome measures by Department, College, and University.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Program Review

Group 1 — Select Final College: All; Select Final Department: All

Head Count by Student Type

		2011	2012	2013	2014	2015
	Cont/Ret	5,586	5,683	5,676	5,879	6,024
	New	2,460	2,433	2,617	2,606	2,766
	Grand Total	8,046	8,116	8,293	8,485	8,790

Underrepresented Minority (URM)

URM	African American	3%	4%	4%	4%	3%
	American Indian	1%	1%	1%	1%	1%
	Hispanic/Latino	19%	22%	26%	29%	31%
	Pacific Islander	0%	0%	0%	0%	0%
	Two or More	4%	4%	4%	4%	4%
	Total	28%	31%	34%	38%	40%
Not URM	Asian American	3%	3%	3%	3%	3%
	Two or More	2%	2%	2%	2%	2%
	White	54%	53%	51%	48%	46%
	Total	58%	58%	56%	54%	51%
Unknown	Other	14%	11%	10%	9%	9%
	Total	14%	11%	10%	9%	9%

Pell Grant

	2011	2012	2013	2014	2015
Non-pell grant recipients	55%	53%	50%	48%	49%
Pell grant recipients	45%	47%	50%	52%	51%

Gender

Female	54%	54%	54%	55%	56%
Male	46%	46%	46%	45%	44%

First Generation

First-gen	44%	46%	49%	51%	52%
Not First-gen	50%	49%	47%	46%	44%
Unknown	6%	4%	4%	3%	4%

Group 2 — Select Final College: All; Select Final Department: All

Head Count by Student Type

		2011	2012	2013	2014	2015
	Cont/Ret	5,586	5,683	5,676	5,879	6,024
	New	2,460	2,433	2,617	2,606	2,766
	Grand Total	8,046	8,116	8,293	8,485	8,790

Underrepresented Minority (URM)

URM	African American	3%	4%	4%	4%	3%
	American Indian	1%	1%	1%	1%	1%
	Hispanic/Latino	19%	22%	26%	29%	31%
	Pacific Islander	0%	0%	0%	0%	0%
	Two or More	4%	4%	4%	4%	4%
	Total	28%	31%	34%	38%	40%
Not URM	Asian American	3%	3%	3%	3%	3%
	Two or More	2%	2%	2%	2%	2%
	White	54%	53%	51%	48%	46%
	Total	58%	58%	56%	54%	51%
Unknown	Other	14%	11%	10%	9%	9%
	Total	14%	11%	10%	9%	9%

Pell Grant

	2011	2012	2013	2014	2015
Non-pell grant recipients	55%	53%	50%	48%	49%
Pell grant recipients	45%	47%	50%	52%	51%

Gender

Female	54%	54%	54%	55%	56%
Male	46%	46%	46%	45%	44%

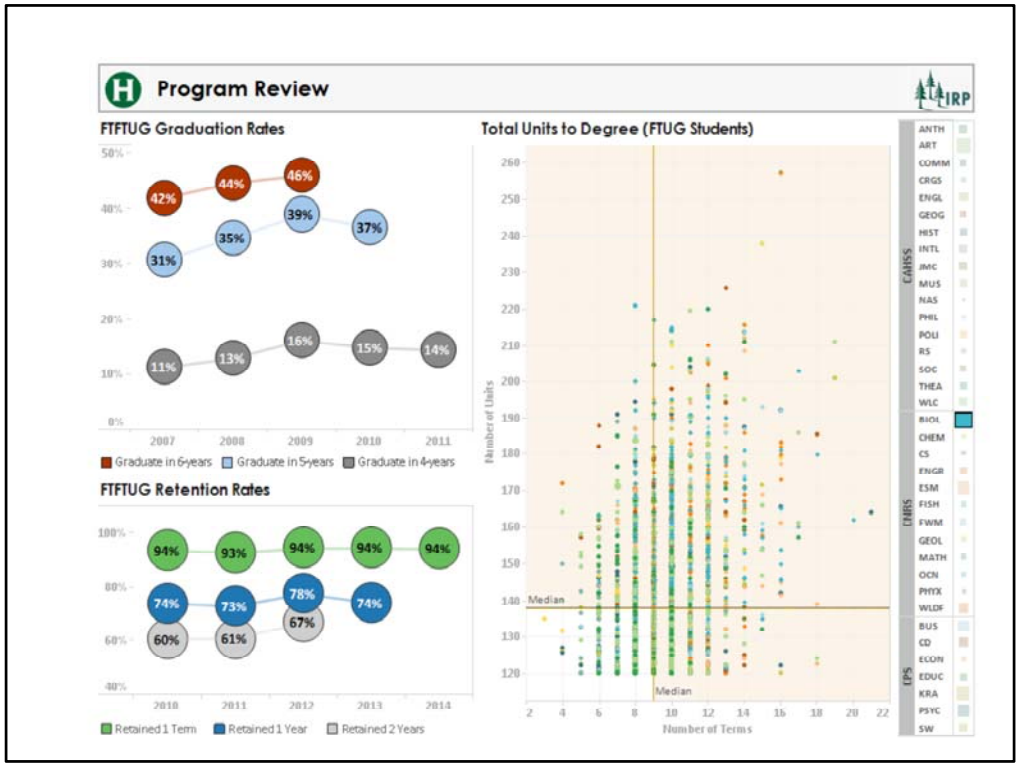
First Generation

First-gen	44%	46%	49%	51%	52%
Not First-gen	50%	49%	47%	46%	44%
Unknown	6%	4%	4%	3%	4%

Page 4

Compare side by side demographics by Department, College, and University.

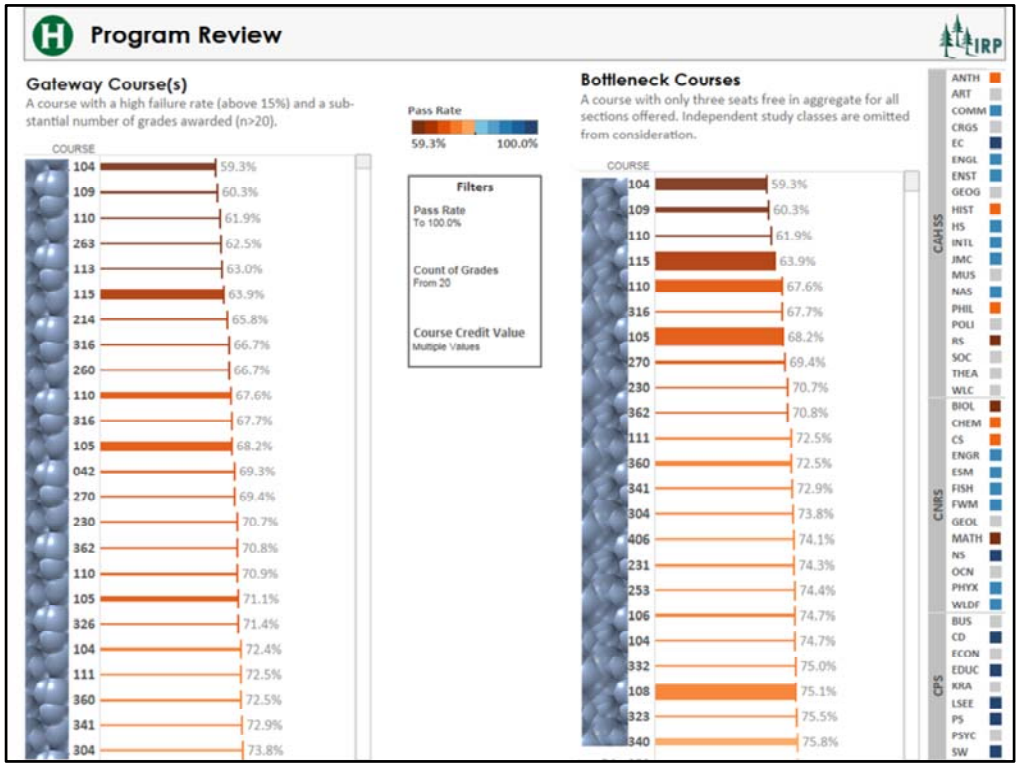
Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Page 5

Exploring outcome measures and leading indicators such as retention and time & units to degree.

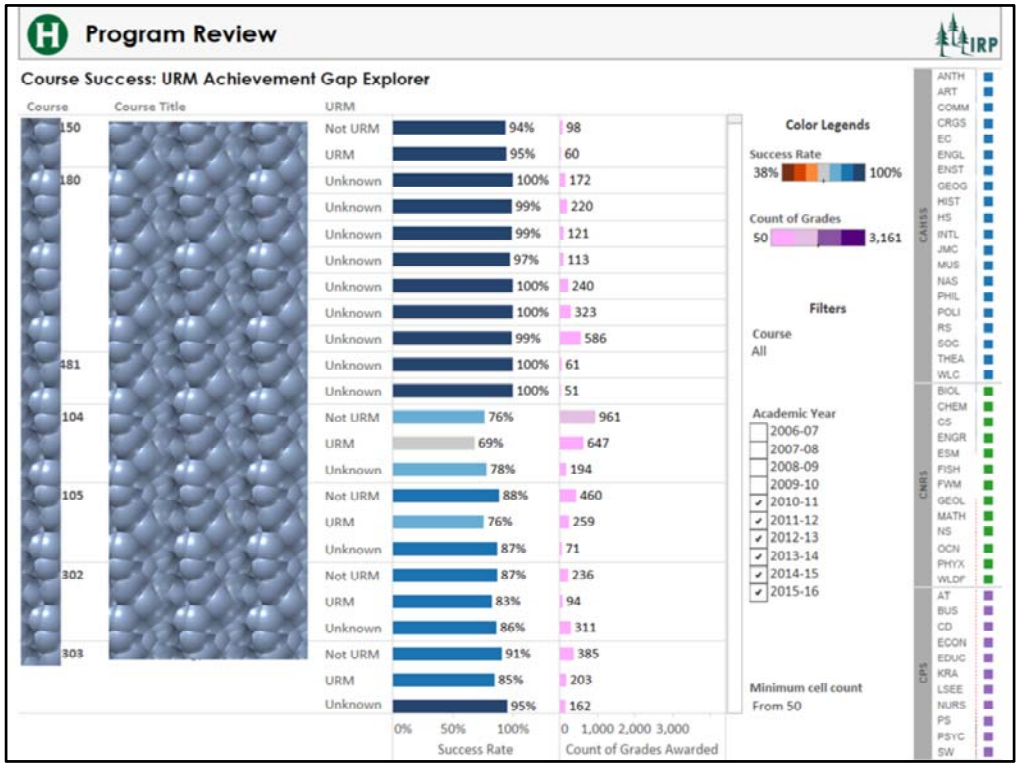
Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Page 6

Explore Gateway and Bottleneck courses side by side. Some data hidden.

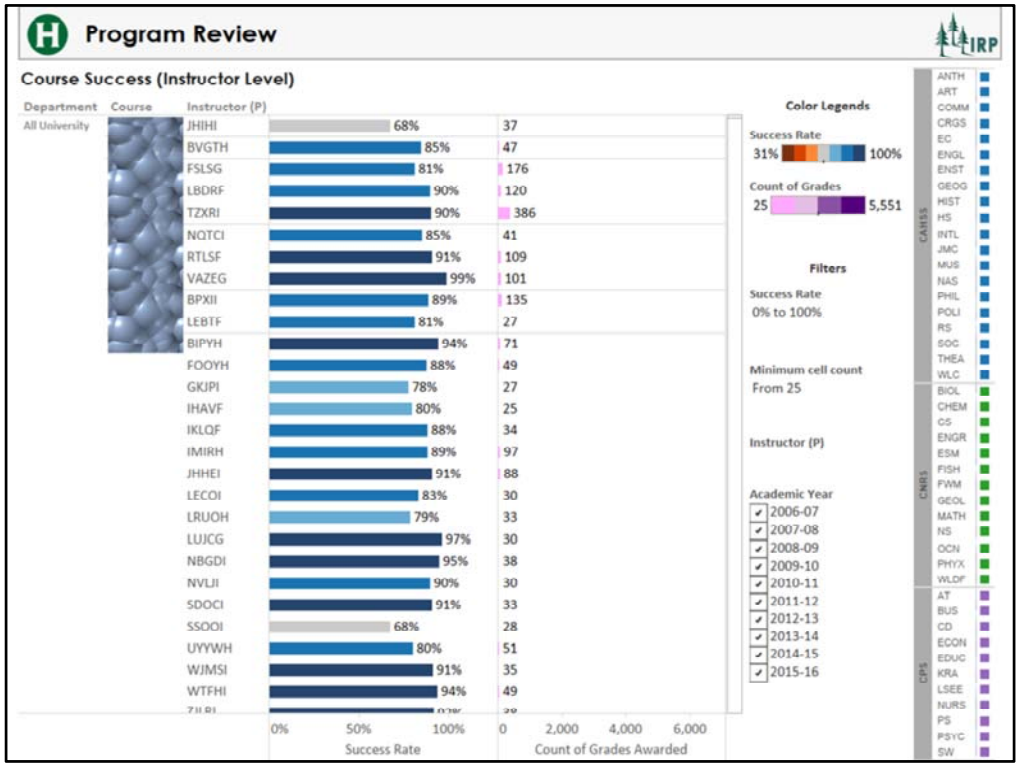
Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Page 7

Explore Underrepresented Minority (URM) achievement gap by course. Some data hidden.

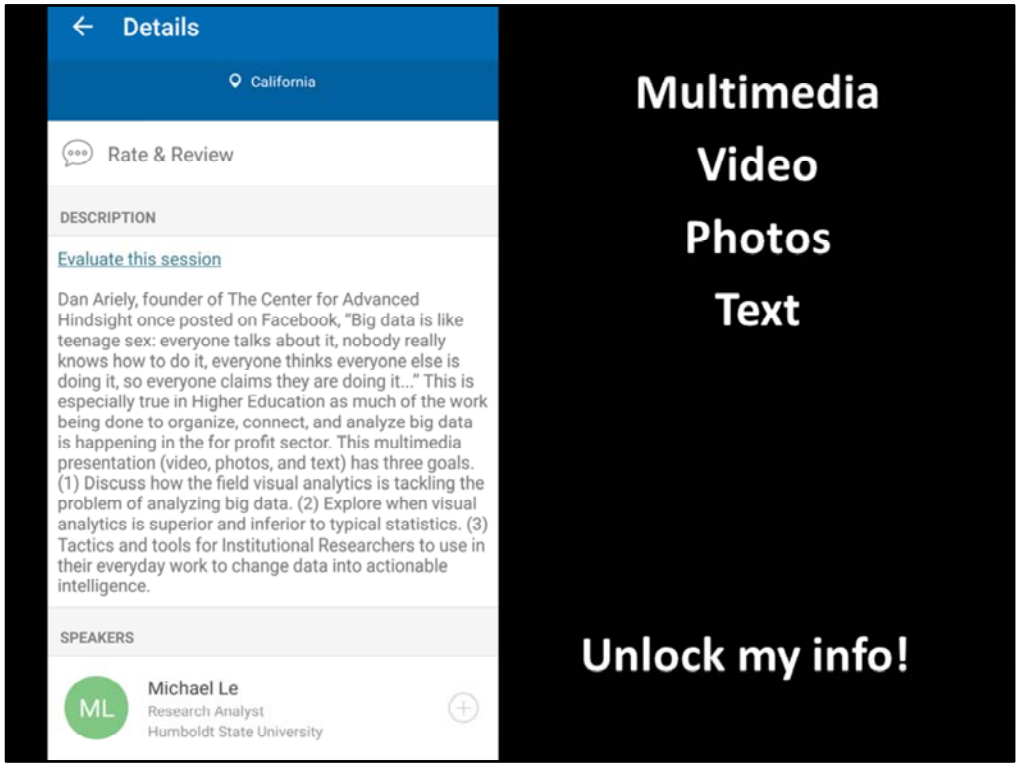
Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Page 8

Explore Courses success by instructor. Instructor names have been give secret codes. Some data hidden.

Le.(2015). Visual Analytics. CAIR, San Francisco, CA



Thank you, before I take questions, please remember to complete your evaluations. This is my first solo CAIR presentation, so your feedback will be very helpful to me! Also remember, I specifically need to know what you thought of the multimedia method I used. Was it helpful or distracting.