Data Visualization Best Practices

Tim Vlamis

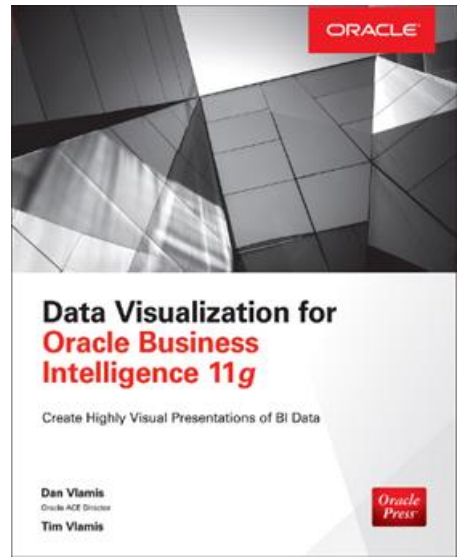
Friday, March 10, 2017

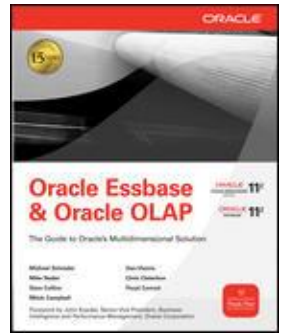
NCOAUG Winter Training Day

@VlamisSoftware

Vlamis Software Solutions

- Vlamis Software founded in 1992 in Kansas City, Missouri
- Developed 200+ Oracle BI and analytics systems
- Specializes in Oracle-based:
  - Enterprise Business Intelligence
  - Data Warehousing
  - Data Mining and Predictive Analytics
  - Data Visualization
- Multiple Oracle ACEs, consultants average 15+ years
- www.vlamis.com (blog, papers, newsletters, services)
- Co-authors of book “Data Visualization for OBI 11g”
- Co-author of book “Oracle Essbase & Oracle OLAP”
- Oracle University Partner
- Oracle Gold Partner





ORACLE EDUCATION RESELLER

ORACLE APPROVED EDUCATION CENTER

ORACLE Gold Partner
Specialized
Oracle Business Intelligence
Foundation Suite 11g

CELEBRATING EXCELLENCE vlamis 25 YEARS SOFTWARE SOLUTIONS

Tim Vlamis

Tim Vlamis – Vice President & Analytics Strategist

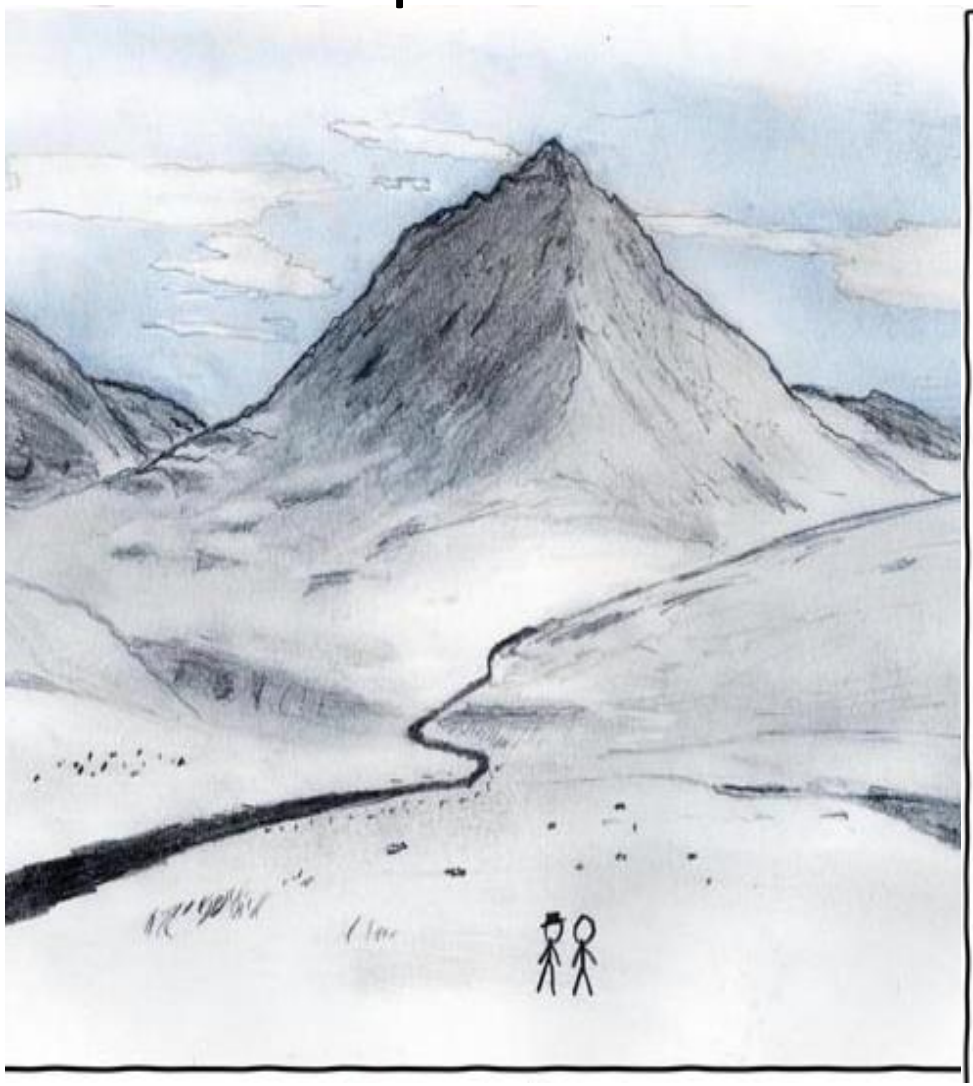
- 30+ years in business modeling and valuation, forecasting, and scenario analyses
- Oracle ACE 
- Instructor for Oracle University's Data Mining Techniques and Oracle R Enterprise Essentials Courses
- Professional Certified Marketer (PCM) from AMA
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University

Agenda

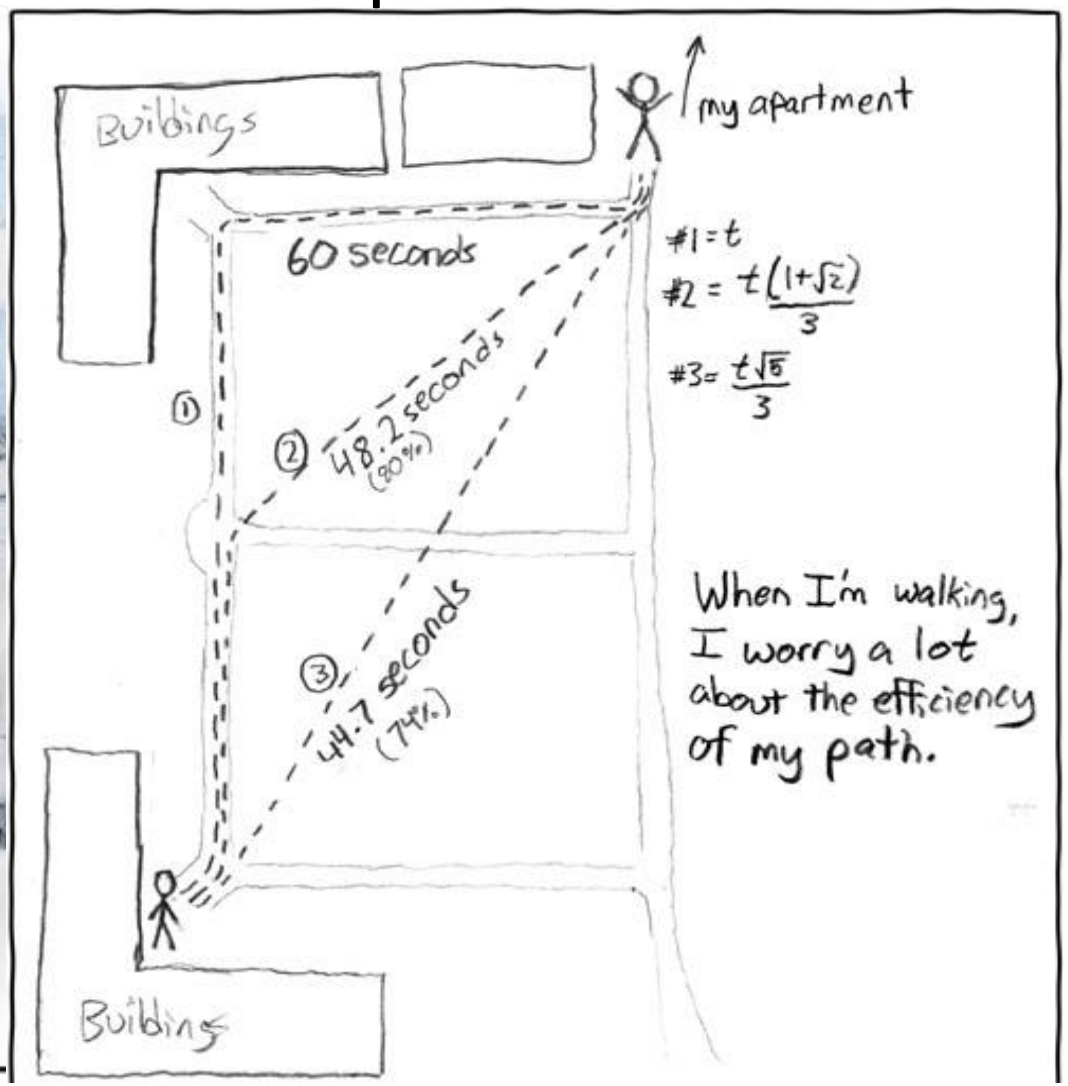
- BI Exploring vs Explanation
- Understanding the foundations of data discovery
- Discovery scenarios
- Discovery frameworks
- Using Visual Analyzer to discover data insights

Main Uses of BI Systems

Exploration

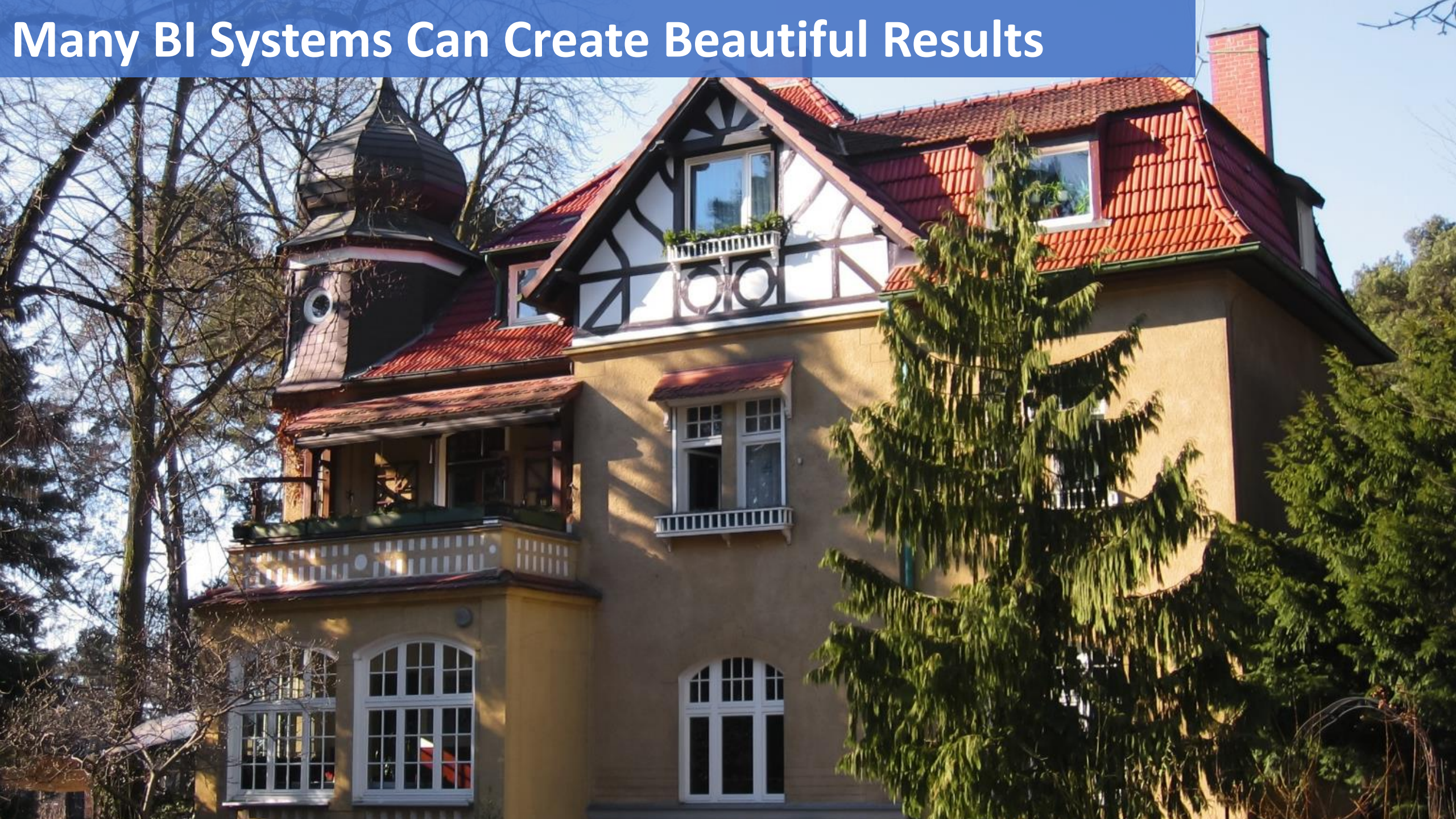


Explanation



xkcd.com

Many BI Systems Can Create Beautiful Results



OBI Operates at a Different Scale

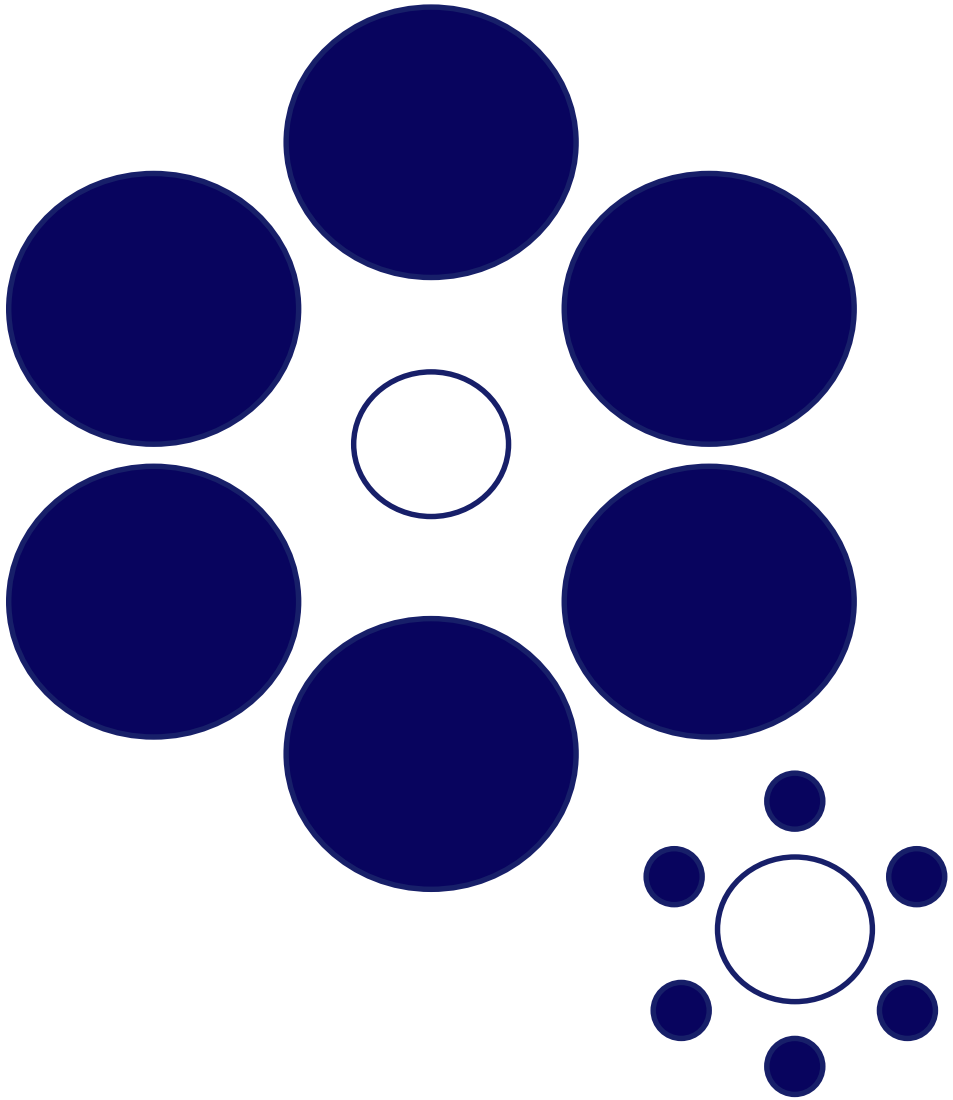

Ingredients → Data Quality & Variety

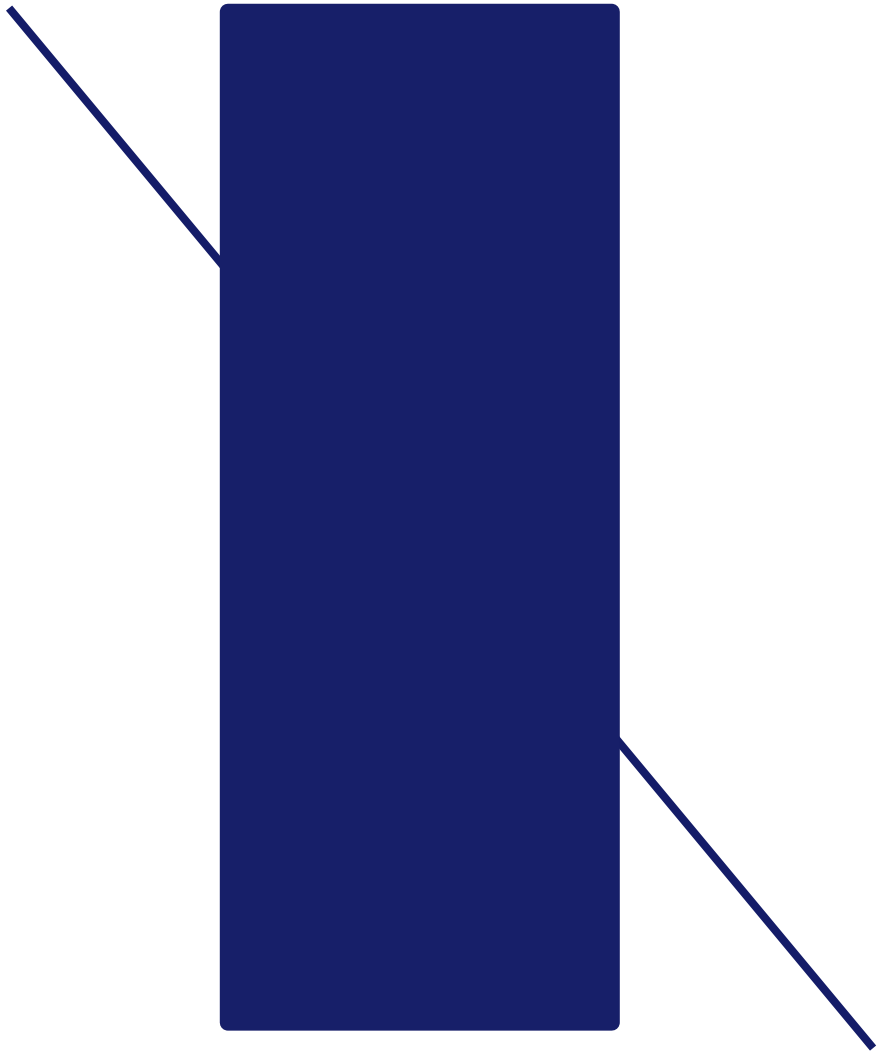
Technique → Data Processing & Prep

Presentation → Data Visualization

The Principles of Human Cognition Should Guide BI Visualization Design

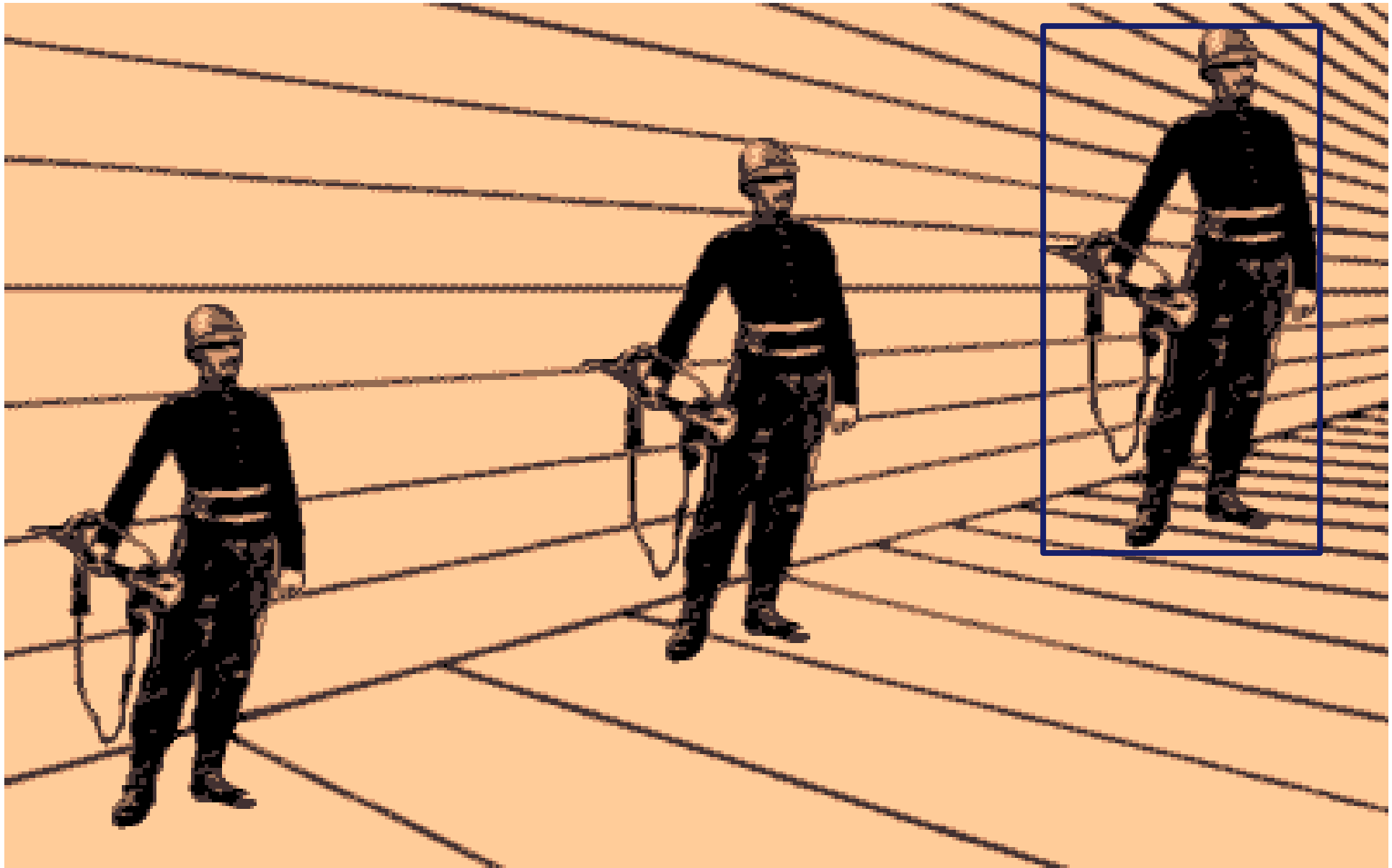
Classic Optical Illusions



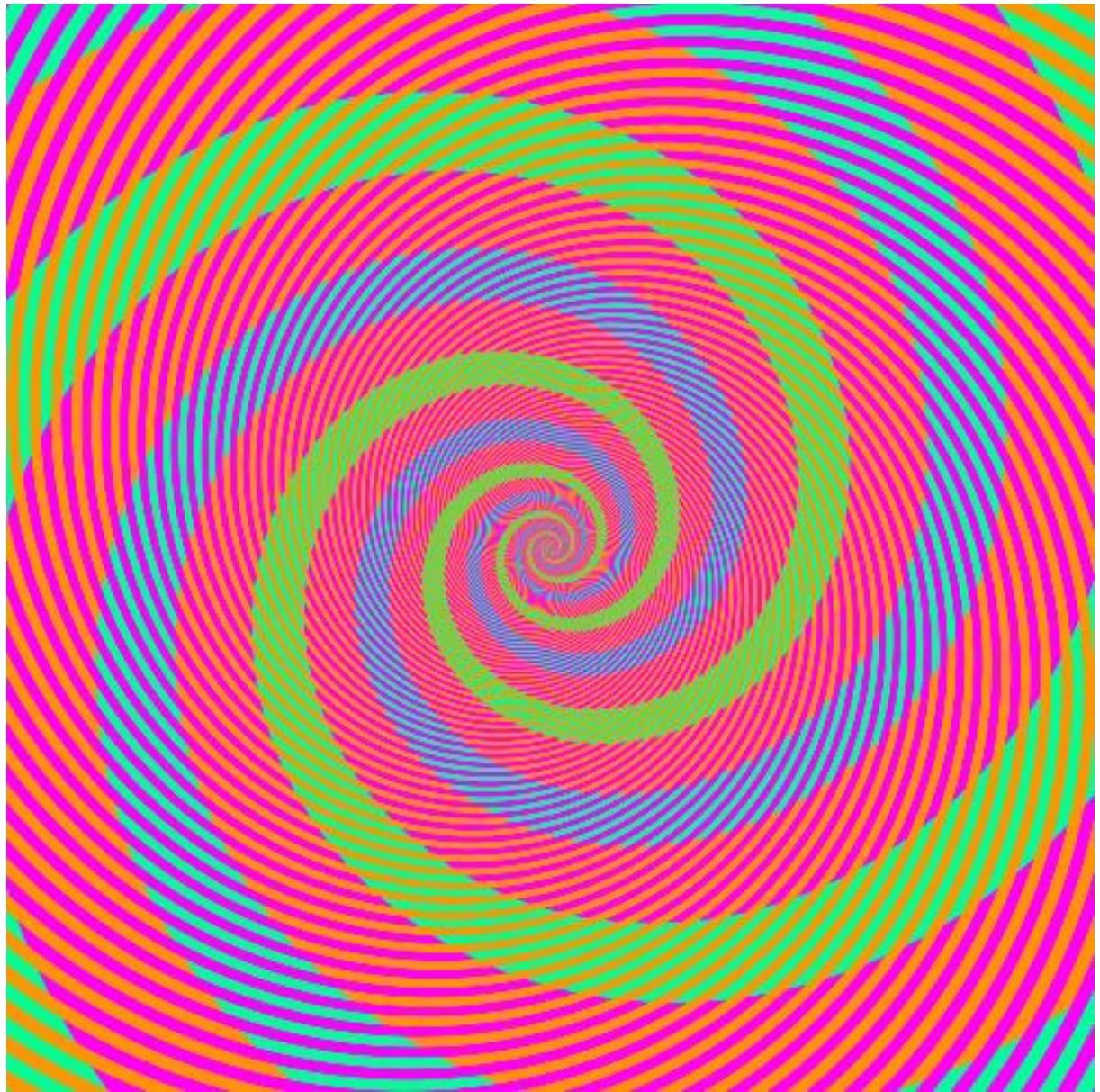


vlamis™
SOFTWARE SOLUTIONS

Which Soldier is Tallest



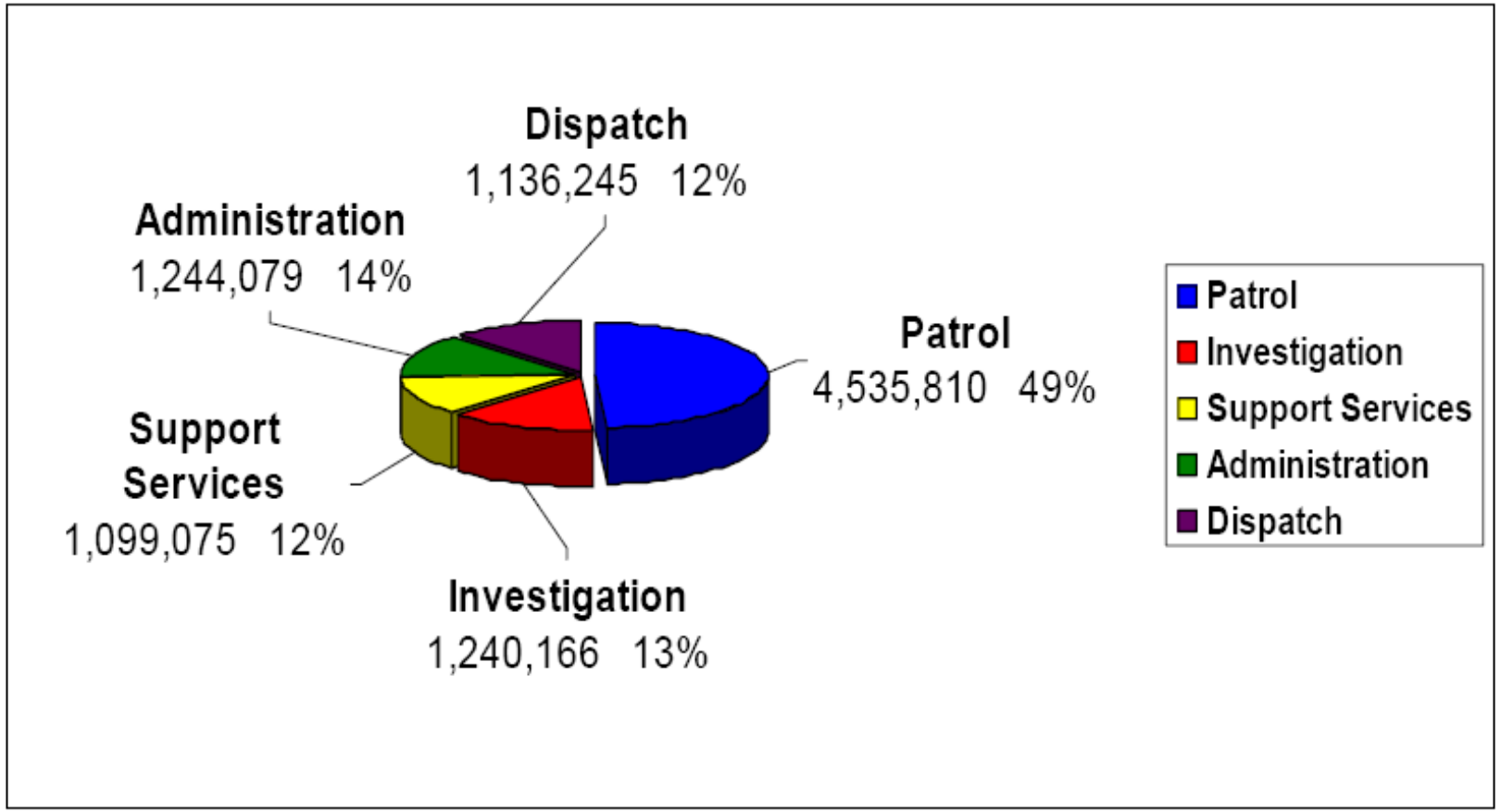
vlamis™
SOFTWARE SOLUTIONS



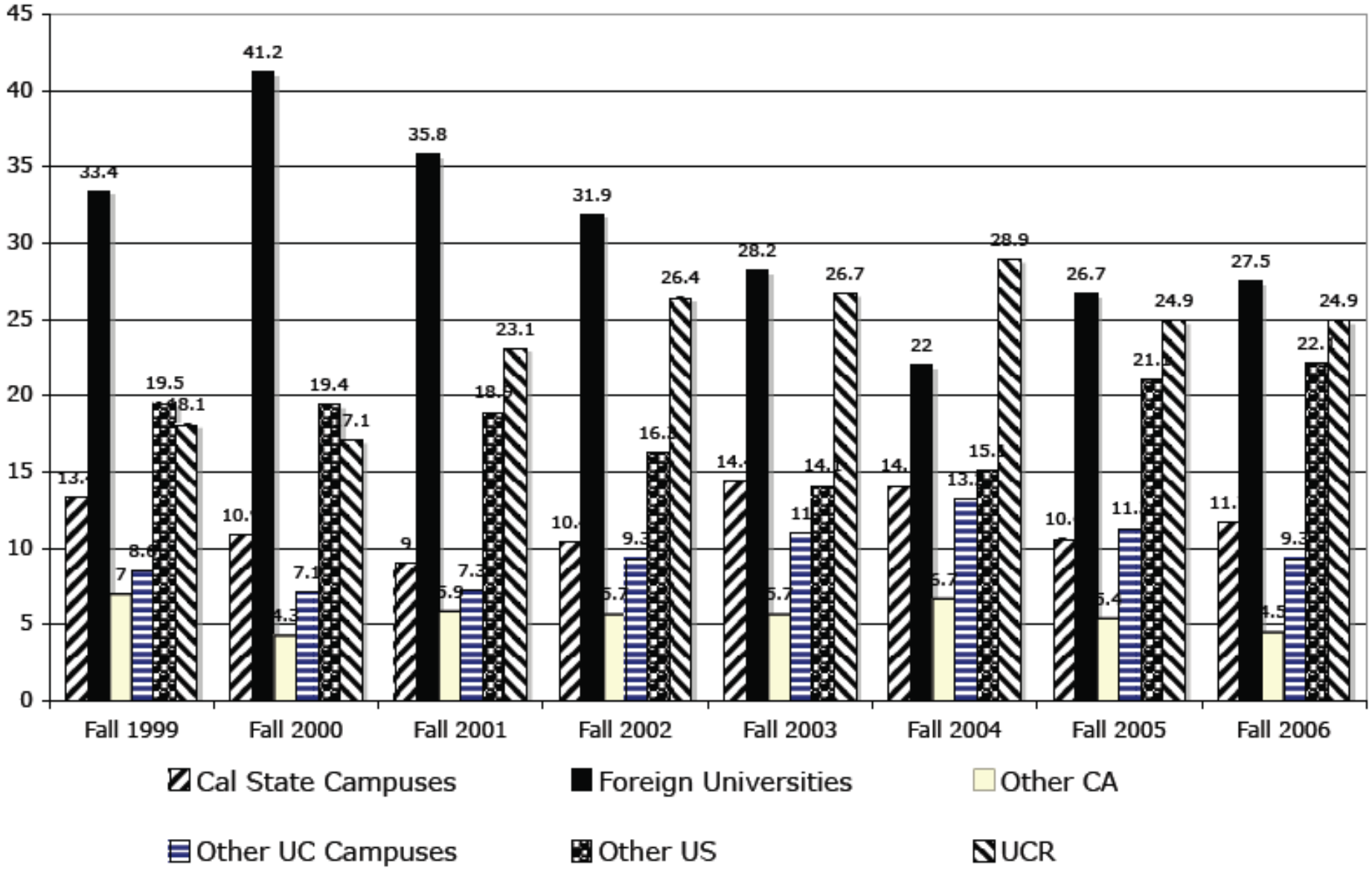
vlamis™
SOFTWARE SOLUTIONS

2004 - 2005 Budget

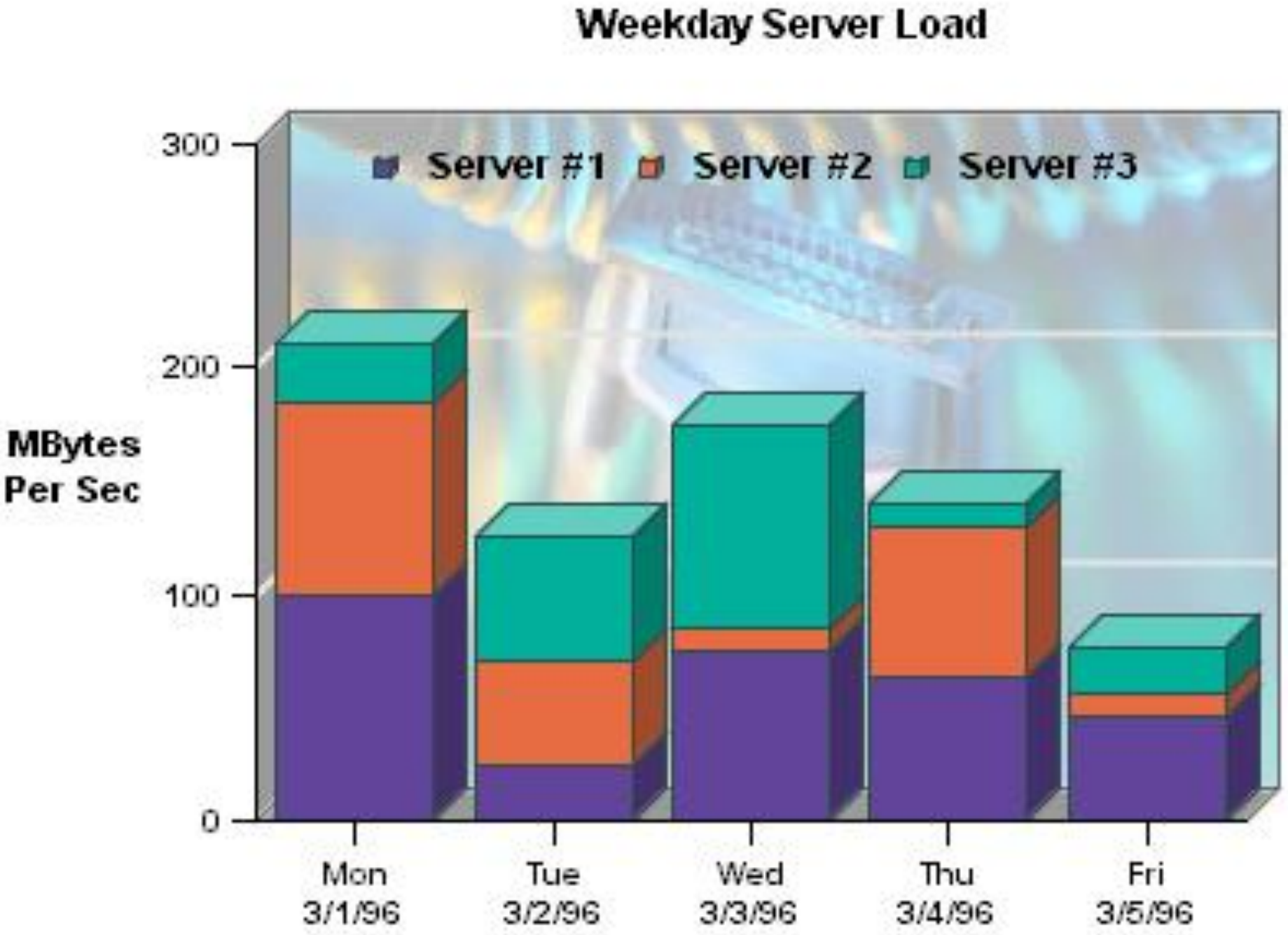
Budget By Division



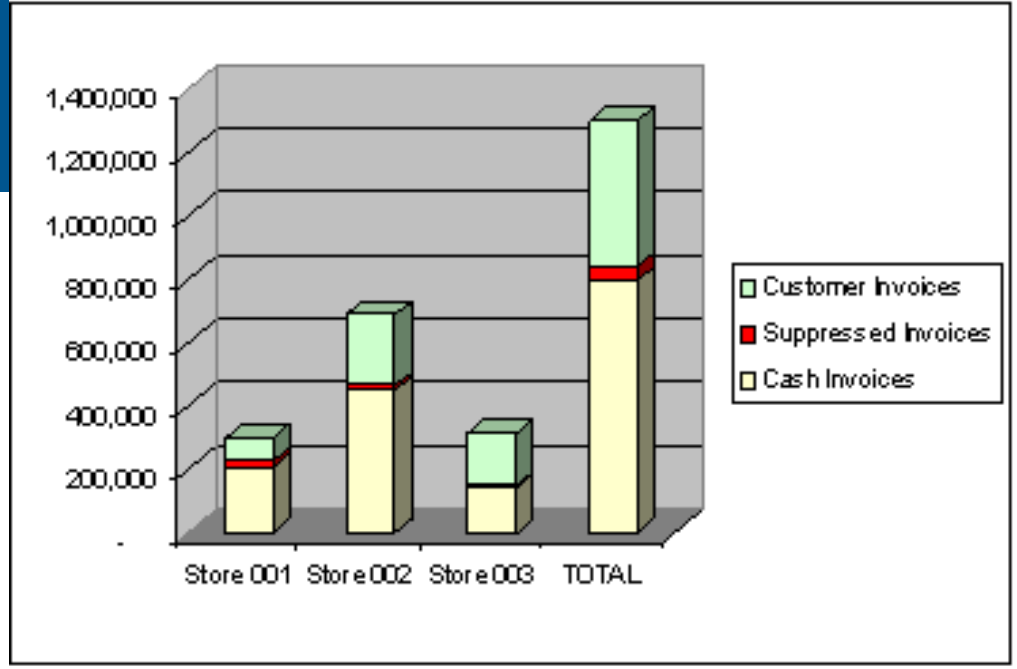
Baccalaureate Degree Institutions of New Graduate Students- Fall Quarters- Percentages from Type of Institution



vlamis™
SOFTWARE SOLUTIONS



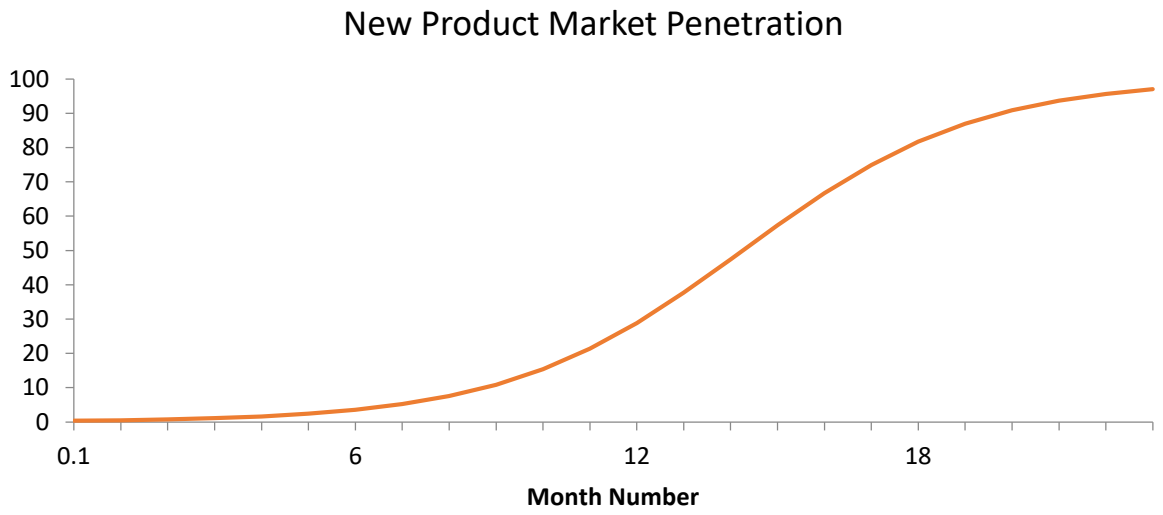
vlamis™
SOFTWARE SOLUTIONS



	Store 001	Store 002	Store 003	TOTAL
Total Invoices	298,943	687,091	313,140	1,299,174
less				
Cash Invoices	207,256	449,064	141,305	797,625
leaves				
Non-cash Invoices	91,687	238,027	171,835	501,549
consisting of				
Suppressed Invoices	18,888	15,527	6,501	40,916
and				
Customer Invoices	72,799	222,500	165,334	460,633
for purchases from				
Suppressed Customer Names	2,123	4,306	870	7,299
and				
Active Customer Names	2,103	14,747	8,342	25,192
which include				
Duplicate Customer Names	70	693	619	1,382
leaving				
Unique Customer Names	2,033	14,054	7,723	23,810
which include				
Bad Addresses	1,055	5,759	2,406	9,220
leaving				
Mailable Customer Names	978	8,295	5,317	14,590

Graphs and Tables

- Graphs and Charts depict visual representations and relationships



- Tables show data organized for lookup of specific, precise values or items.

Order Type	No of Orders	Sales	Billed Quantity	Actual Unit Price
Express	13,980	$14,027,034	1,117,199	$12.56
Secure	29,347	$28,513,745	2,326,540	$12.26
Standard	27,673	$27,459,221	2,213,482	$12.41
Grand Total	**71,000**	**$70,000,000**	**5,657,221**	**$12.37**

vlamis™
SOFTWARE SOLUTIONS

Characteristics of Tables

- Can present data at drastically different scales.
- Can present very different data types simultaneously.
- Can repeat and include multiple sets of the same data values.
- Are extraordinarily dense and include numerous data relationships without direct distortion of the data itself.
- Tables can present “federated” data from different sources in a single simultaneous view.

Keys to Effective Tables

- Prefer smaller tables
- Words are important
  - Enable roll overs for meta data for commonly used tables
  - Write informative titles for tables and column head descriptions
- Make tables clean and easy to read
  - Eliminate unnecessary gridlines
  - Use space (padding) to create groups of data
  - Left justify text cells and Right justify numerical cells
- Make numbers easy to read and understand
  - Judiciously use conditional formatting
  - Avoid putting text in color
  - Align the decimal point for numerical cells
  - Use symbols to denote units of measure (%, $, etc.)
- Enable column and row sorting
- Avoid scrolling (if possible)
- Be transparent about data selection

Keys to Effective Tables

Year 2010

Product Type	Company	Sales						
		Active Singles	Baby Boomers	Others	Rural based	Seniors	Students	Urban based
Accessories	Genmind Corp	$95,916	$29,746	$23,710	$40,947	$60,397	$59,891	$77,722
	Stockplus Inc.	$128,470	$29,693	$38,455	$68,506	$100,349	$120,508	$111,572
	Tescare Ltd.	$104,461	$35,374	$27,900	$56,392	$96,501	$121,121	$93,280
Accessories Total		**$328,847**	**$94,813**	**$90,064**	**$165,845**	**$257,247**	**$301,520**	**$282,574**
Audio	Genmind Corp	$168,612	$50,236	$21,842	$74,952	$126,754	$133,788	$124,072
	Stockplus Inc.	$215,921	$42,336	$55,632	$124,469	$149,511	$169,330	$144,029
	Tescare Ltd.	$173,022	$61,713	$30,048	$102,717	$162,078	$202,451	$161,995
Audio Total		**$557,555**	**$154,285**	**$107,522**	**$302,137**	**$438,343**	**$505,569**	**$430,096**
Camera	Genmind Corp	$154,930	$50,453	$23,935	$73,360	$129,189	$143,608	$136,459
	Stockplus Inc.	$189,520	$45,571	$57,449	$88,445	$154,237	$181,047	$162,000
	Tescare Ltd.	$182,757	$83,650	$45,512	$89,213	$140,187	$208,441	$151,215
Camera Total		**$527,207**	**$179,675**	**$126,895**	**$251,019**	**$423,613**	**$533,096**	**$449,674**
Cell Phones	Genmind Corp	$120,376	$40,799	$24,293	$61,451	$82,200	$103,754	$97,480
	Stockplus Inc.	$161,238	$47,570	$37,670	$71,548	$129,511	$133,459	$144,812
	Tescare Ltd.	$157,717	$50,948	$30,873	$79,242	$130,167	$164,272	$116,630
Cell Phones Total		**$439,331**	**$139,317**	**$92,837**	**$212,241**	**$341,879**	**$401,484**	**$358,921**
Fixed	Genmind Corp	$144,814	$35,190	$20,000	$94,115	$128,411	$152,767	$138,280
	Stockplus Inc.	$234,518	$56,263	$53,554	$109,985	$160,065	$238,484	$180,872
	Tescare Ltd.	$197,073	$57,671	$50,893	$121,302	$170,018	$173,601	$177,137

vlamis™
SOFTWARE SOLUTIONS

Pivot Table “Needs” Sentence

I want to see fact/measure (specifies cell values)

***by** dimension and dimension (defines rows)*

***across** dimension and dimension (defines columns).*

Year 2010

		Sales						
Product Type	Company	Active Singles	Baby Boomers	Others	Rural based	Seniors	Students	Urban based
Accessories	Genmind Corp	$95,916	$29,746	$23,710	$40,947	$60,397	$59,891	$77,722
	Stockplus Inc.	$128,470	$29,693	$38,455	$68,506	$100,349	$120,508	$111,572
	Tescare Ltd.	$104,461	$35,374	$27,900	$56,392	$96,501	$121,121	$93,280
Accessories Total		**$328,847**	**$94,813**	**$90,064**	**$165,845**	**$257,247**	**$301,520**	**$282,574**
Audio	Genmind Corp	$168,612	$50,236	$21,842	$74,952	$126,754	$133,788	$124,072
	Stockplus Inc.	$215,921	$42,336	$55,632	$124,469	$149,511	$169,330	$144,029
	Tescare Ltd.	$173,022	$61,713	$30,048	$102,717	$162,078	$202,451	$161,995
Audio Total		**$557,555**	**$154,285**	**$107,522**	**$302,137**	**$438,343**	**$505,569**	**$430,096**
Camera	Genmind Corp	$154,930	$50,453	$23,935	$73,360	$129,189	$143,608	$136,459
	Stockplus Inc.	$189,520	$45,571	$57,449	$88,445	$154,237	$181,047	$162,000
	Tescare Ltd.	$182,757	$83,650	$45,512	$89,213	$140,187	$208,441	$151,215
Camera Total		**$527,207**	**$179,675**	**$126,895**	**$251,019**	**$423,613**	**$533,096**	**$449,674**
Cell Phones	Genmind Corp	$120,376	$40,799	$24,293	$61,451	$82,200	$103,754	$97,480
	Stockplus Inc.	$161,238	$47,570	$37,670	$71,548	$129,511	$133,459	$144,812
	Tescare Ltd.	$157,717	$50,948	$30,873	$79,242	$130,167	$164,272	$116,630
Cell Phones Total		**$439,331**	**$139,317**	**$92,837**	**$212,241**	**$341,879**	**$401,484**	**$358,921**
Fixed	Genmind Corp	$144,814	$35,190	$20,000	$94,115	$128,411	$152,767	$138,280
	Stockplus Inc.	$234,518	$56,263	$53,554	$109,985	$160,065	$238,484	$180,872
	Tescare Ltd.	$197,073	$57,671	$50,893	$121,302	$170,018	$173,601	$177,137

vlamis™
SOFTWARE SOLUTIONS

Pivot Table “Needs” Sentence

I want to see Sales (specifies cell values)

***by** Product Type and Company (defines rows)*

***across** Market Segments (defines columns).*

Year 2010

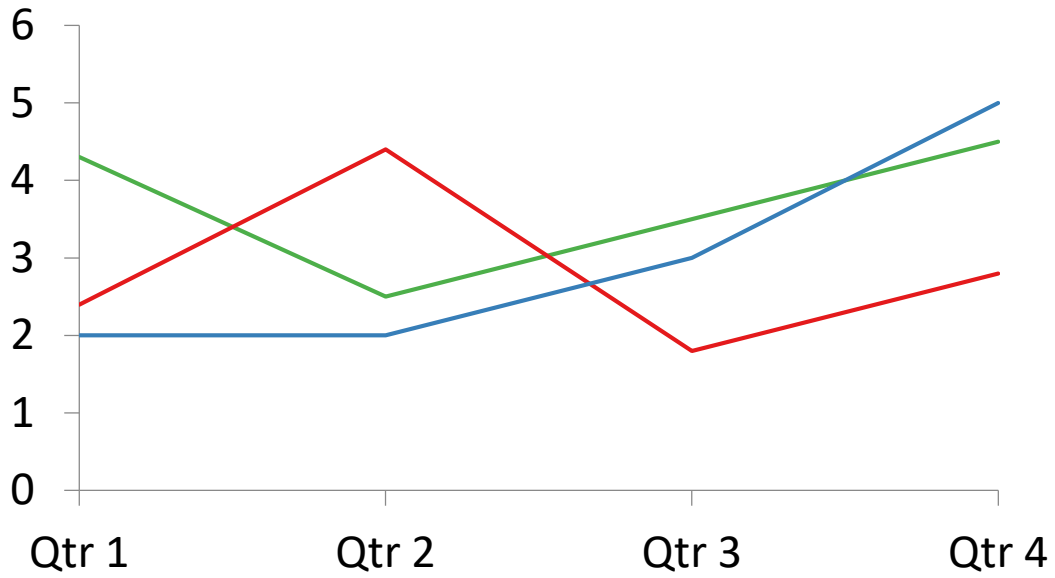
		Sales						
Product Type	Company	Active Singles	Baby Boomers	Others	Rural based	Seniors	Students	Urban based
Accessories	Genmind Corp	$95,916	$29,746	$23,710	$40,947	$60,397	$59,891	$77,722
	Stockplus Inc.	$128,470	$29,693	$38,455	$68,506	$100,349	$120,508	$111,572
	Tescare Ltd.	$104,461	$35,374	$27,900	$56,392	$96,501	$121,121	$93,280
Accessories Total		**$328,847**	**$94,813**	**$90,064**	**$165,845**	**$257,247**	**$301,520**	**$282,574**
Audio	Genmind Corp	$168,612	$50,236	$21,842	$74,952	$126,754	$133,788	$124,072
	Stockplus Inc.	$215,921	$42,336	$55,632	$124,469	$149,511	$169,330	$144,029
	Tescare Ltd.	$173,022	$61,713	$30,048	$102,717	$162,078	$202,451	$161,995
Audio Total		**$557,555**	**$154,285**	**$107,522**	**$302,137**	**$438,343**	**$505,569**	**$430,096**
Camera	Genmind Corp	$154,930	$50,453	$23,935	$73,360	$129,189	$143,608	$136,459
	Stockplus Inc.	$189,520	$45,571	$57,449	$88,445	$154,237	$181,047	$162,000
	Tescare Ltd.	$182,757	$83,650	$45,512	$89,213	$140,187	$208,441	$151,215
Camera Total		**$527,207**	**$179,675**	**$126,895**	**$251,019**	**$423,613**	**$533,096**	**$449,674**
Cell Phones	Genmind Corp	$120,376	$40,799	$24,293	$61,451	$82,200	$103,754	$97,480
	Stockplus Inc.	$161,238	$47,570	$37,670	$71,548	$129,511	$133,459	$144,812
	Tescare Ltd.	$157,717	$50,948	$30,873	$79,242	$130,167	$164,272	$116,630
Cell Phones Total		**$439,331**	**$139,317**	**$92,837**	**$212,241**	**$341,879**	**$401,484**	**$358,921**
Fixed	Genmind Corp	$144,814	$35,190	$20,000	$94,115	$128,411	$152,767	$138,280
	Stockplus Inc.	$234,518	$56,263	$53,554	$109,985	$160,065	$238,484	$180,872
	Tescare Ltd.	$197,073	$57,671	$50,893	$121,302	$170,018	$173,601	$177,137

vlamis™
SOFTWARE SOLUTIONS

7 Keys to Effective Graphs

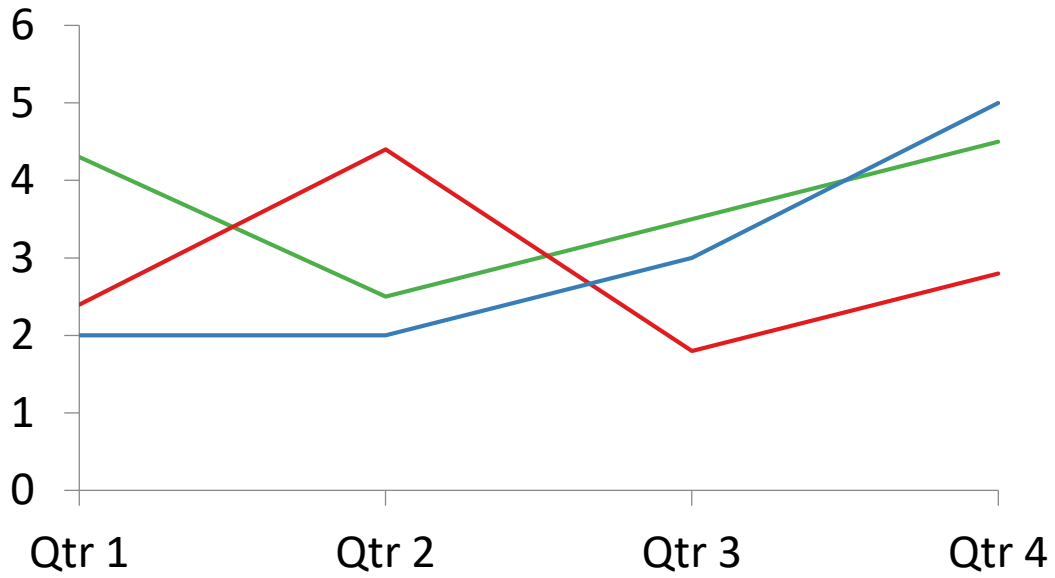
- Do not use 3-D effects.
- Avoid “stop light” color palette.
- Prefer pastel color palettes and avoid bright colors.
- Eliminate gridlines, drop shadows, and other graphics.
- Enable interaction for “exploration” graphs.
- Prioritize a single message for “explanation” graphs.
- Above all else, show the data!

Line Graph



- Show a pattern or progression over a continuous range.
- Can be valued within a range to highlight a particular pattern (careful!).
- Maintain a rectangular shape close to golden proportion.
- Use scale marker lines and ranges for context.

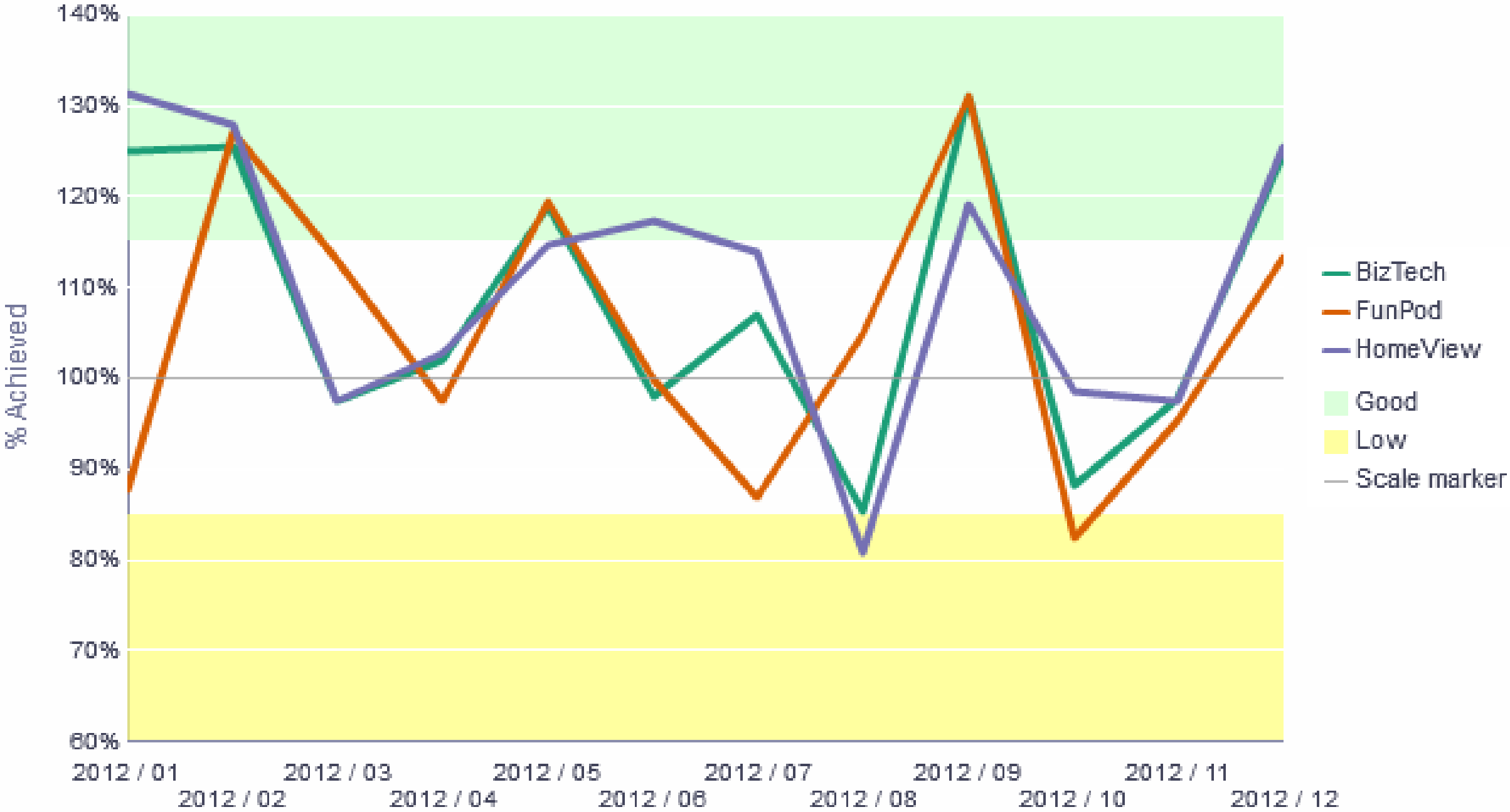
Line Graph



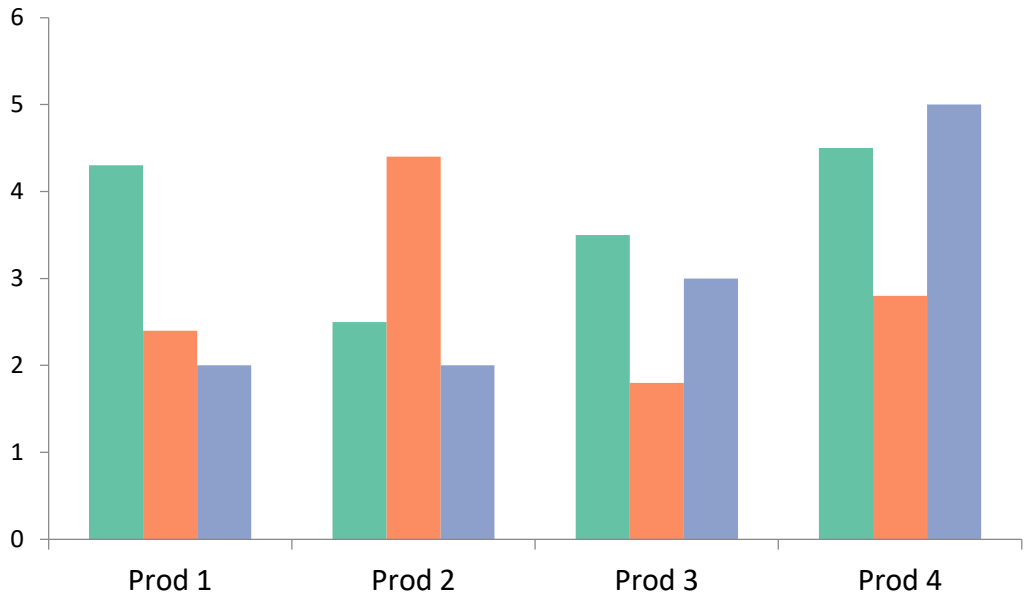
- Use darker versions of standard colors.
- Eliminate grid lines.
- Use zoom function for detailed line graphs.
- Choose curved lines to smooth overall shape.
- Choose stepped lines to emphasize point transitions.

vlamis™
SOFTWARE SOLUTIONS

Target Revenue % by Brand for 2012

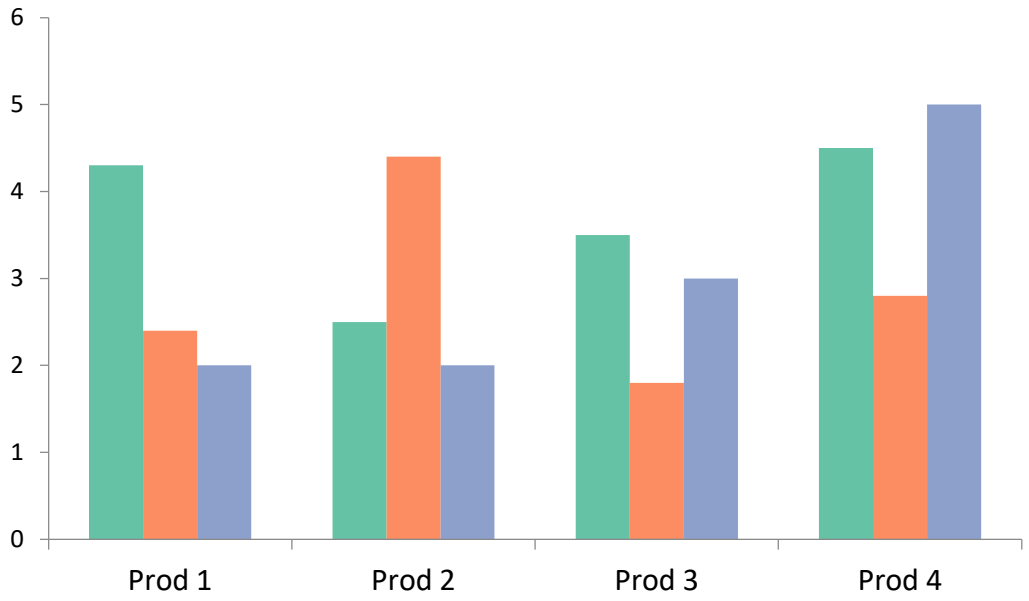


Bar Graphs



- Show nominal data values in comparison to one another.
- Start with zero.
- If use a logarithmic scale, clearly notate.
- Think through sort order carefully.

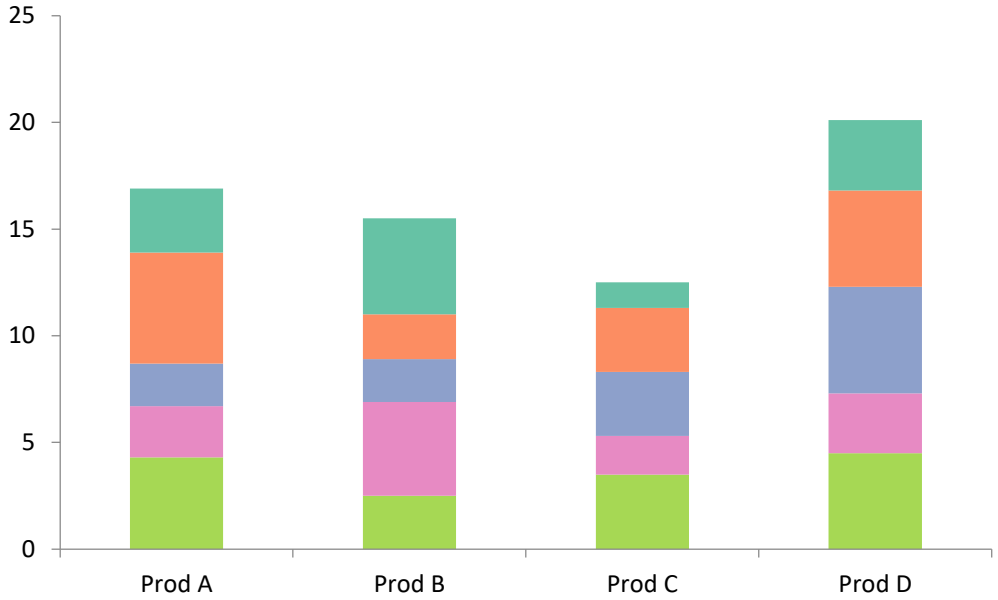
Bar Graphs



- Add data labels as interactive rollover.
- Balance colors.
- If change is most important, graph change.

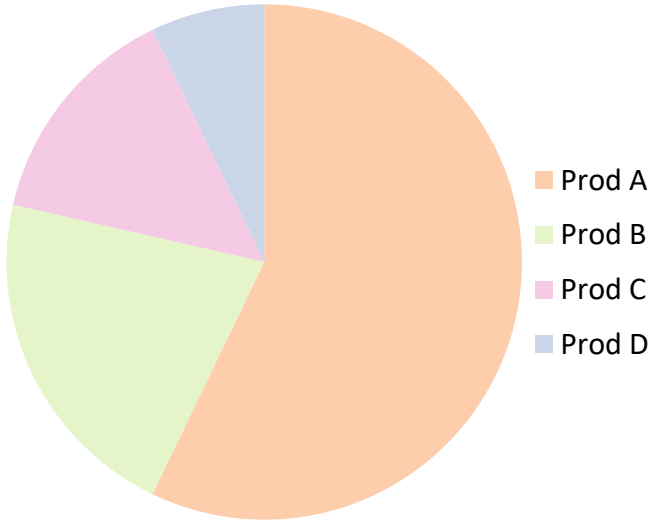
vlamis™
SOFTWARE SOLUTIONS

Stacked Bar Chart



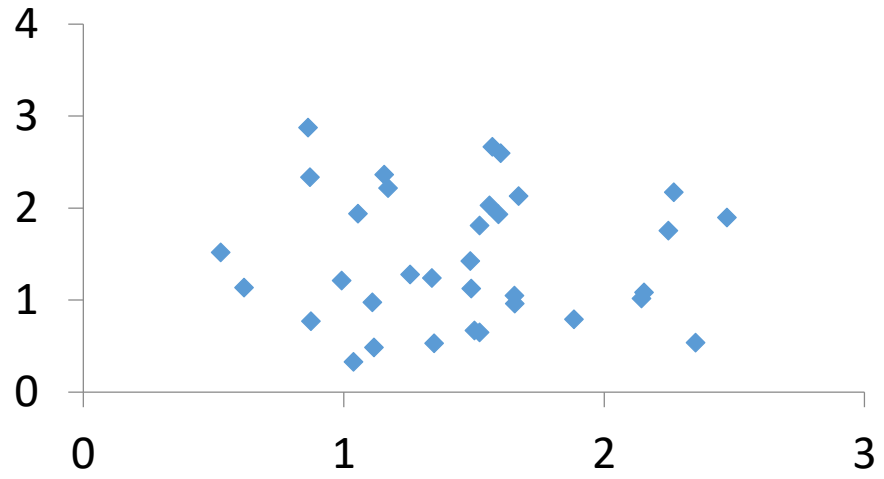
- Somewhat confusing, not great for representing change.
- Total is most clearly represented number.
- Typically stack with largest values on the bottom.
- Single scale can make for interesting intra-bar comparisons.

Pie Charts



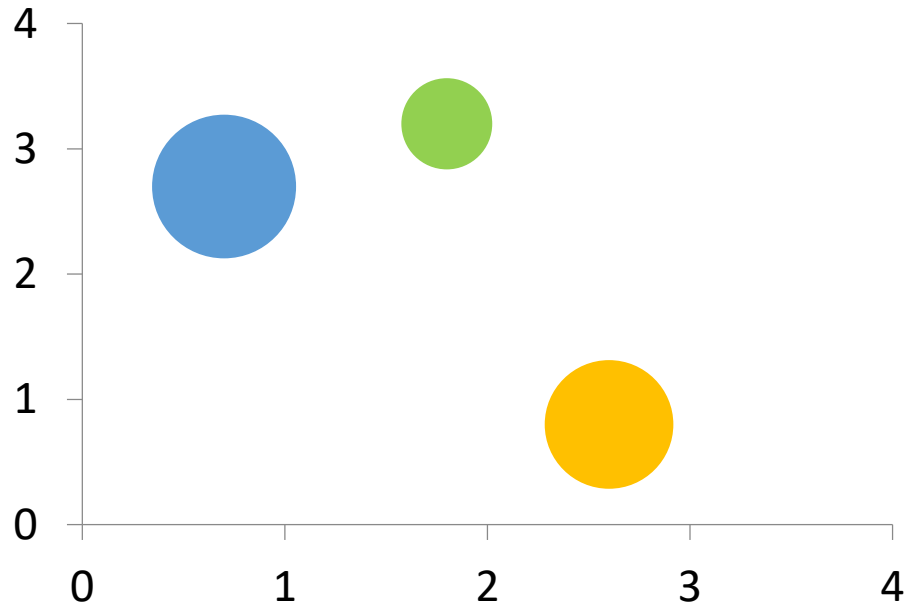
- Typically used for showing parts of whole by percentage.
- Not great for piece to piece comparisons.
- Limit number of pieces.
- Can be interesting to show many pies together if significant differences exist.
- Stephen Few hates them.
- Do not use 3-D.

Scatter Plot



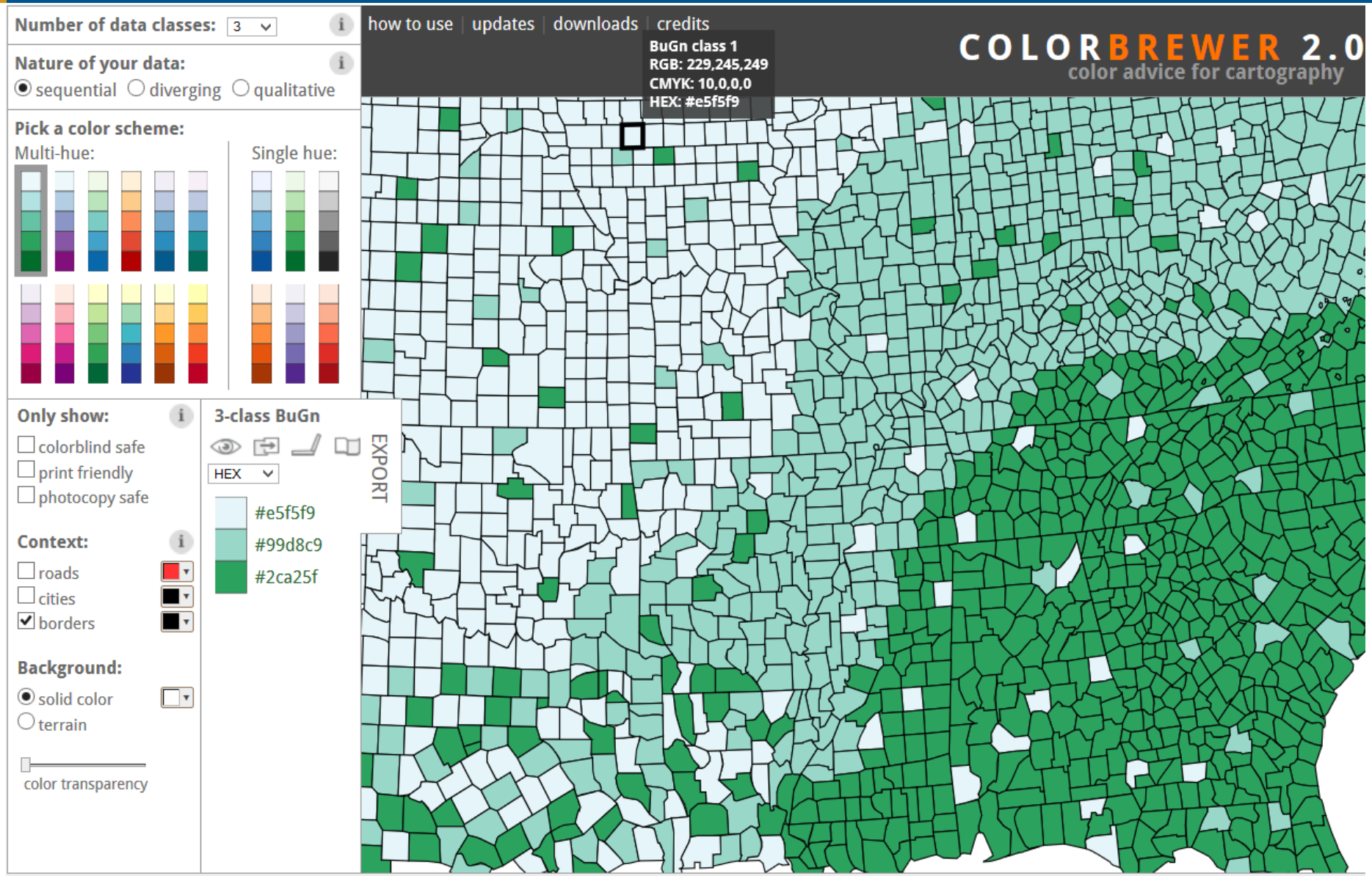
- Shows single data points at the intersection of two values.
- Often depict a large number of discrete data points (hundreds or thousands).
- Useful comparisons of two variables.
- Trend lines are often added.
- Clearly notate if use logarithmic scale(s).

Bubble Chart



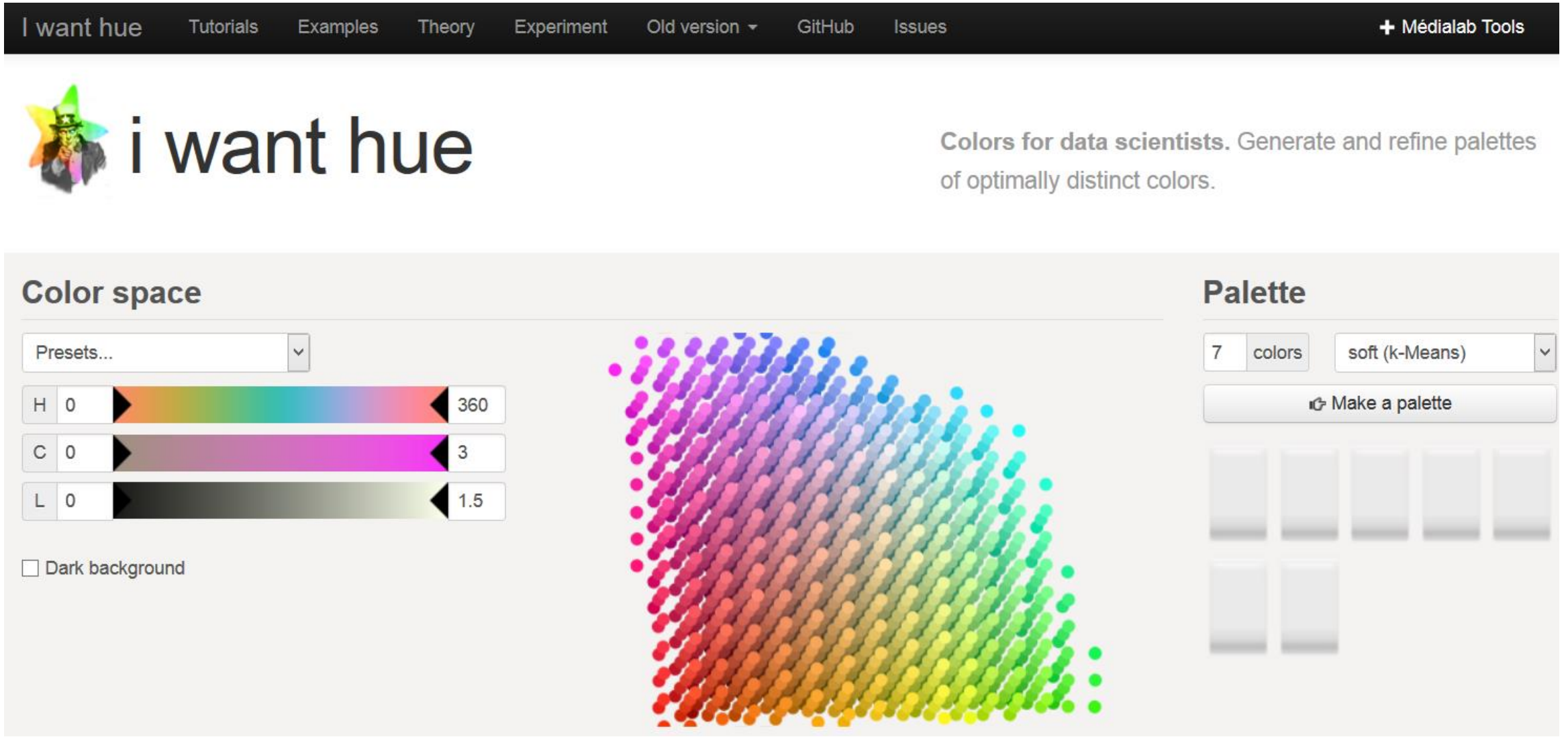
- Special type of scatter plot.
- Size of bubble is related to a third variable.
- Color is related to a fourth variable.
- Reduces number of points that can be depicted.
- Best for depicting approximate values and comparisons.

ColorBrewer2.org



vlamis™
SOFTWARE SOLUTIONS

i want hue



vlamis™
SOFTWARE SOLUTIONS

Dashboard Definition

A Dashboard is a visual presentation of current summary information needed to manage and guide an organization or activity.

BI Dashboards are Different

- No mechanical systems needed to move indicators.
- Decisions are not typically made on a second-to-second basis.
- BI dashboards are not primarily single situation or single person devices.

BI Dashboards

- Role-based.
- Data selection and filtering are extremely important.
- Dashboards support evidenced-based decision making.
- Shared understanding of business situation is a key benefit.
- Content may be individualized.
- Design should be standardized.

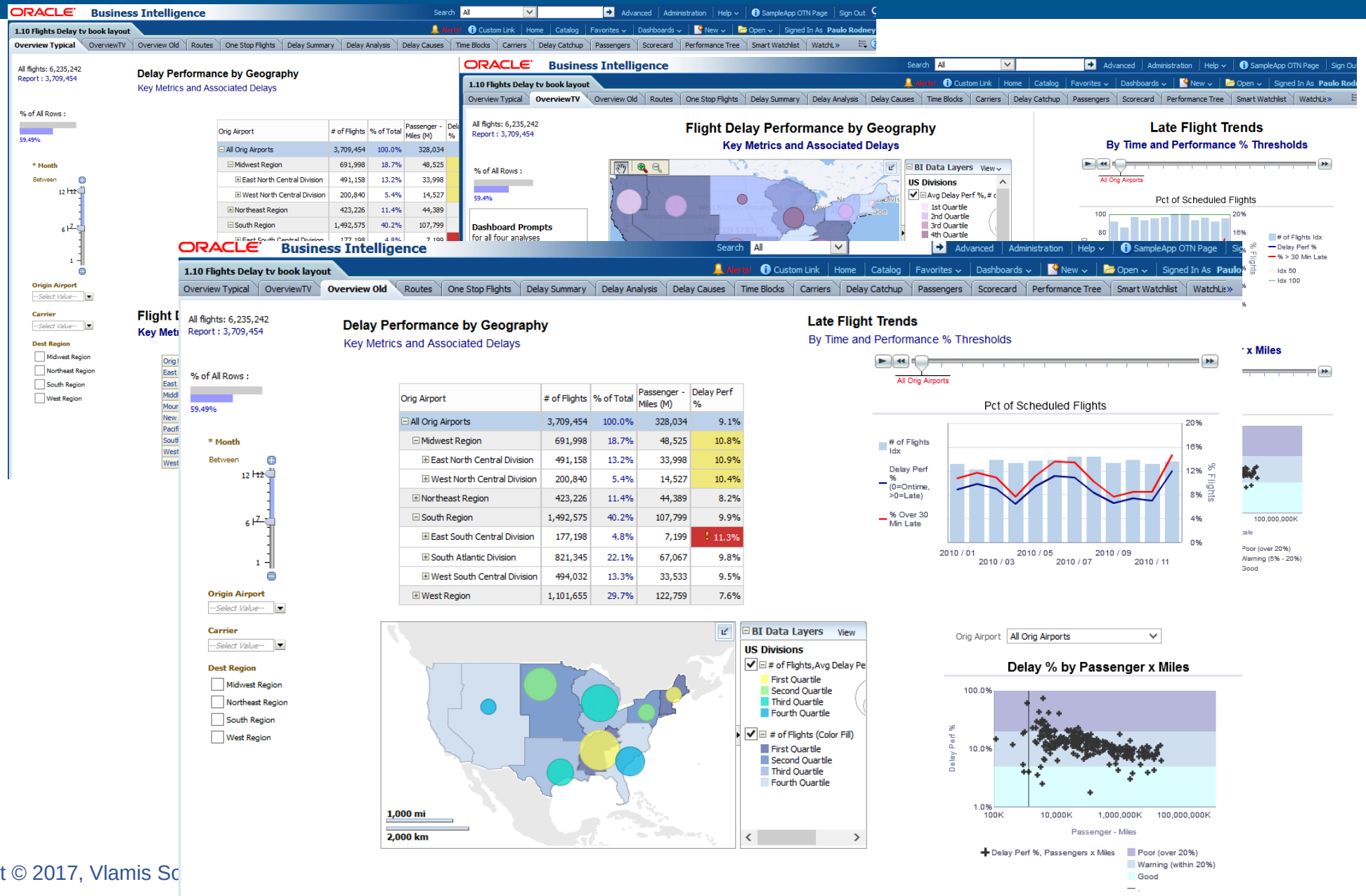
OBIEE Dashboard Overview

- Designed with columns and sections (containers).
- Presentation server is often separate from BI server.
- Dashboards are web-based and are viewed with browsers.
- HTML, XML, and Java coding skills are useful, but not required.

Dashboard Principles

- Promote user interactivity
  - Prompts
  - View and column selectors
  - Hierarchical column drills
  - Column sorts
  - Guided navigation and action links
- Promote data transparency
  - Prompts
  - Filter views
  - Narrative views
  - Master detail linking
- Establish design guidelines for consistency

OBIEE Demo Content from Chap 1



All flights: 6,235,242
Report : 3,709,454

% of All Rows :
59.49%

* Month
Between
12 12
6 7
1

Origin Airport
--Select Value--

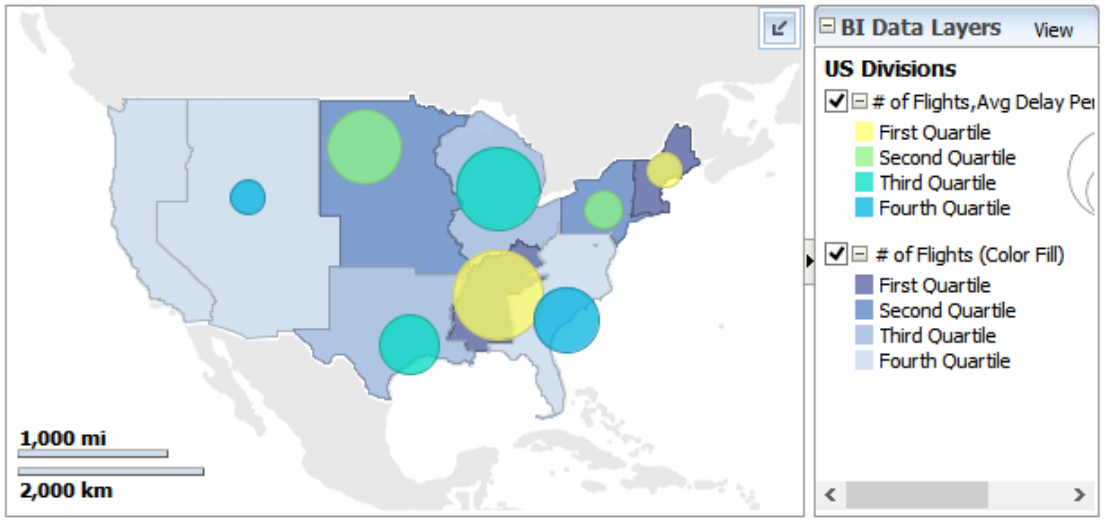
Carrier
--Select Value--

Dest Region
- Midwest Region
- Northeast Region
- South Region
- West Region

Delay Performance by Geography

Key Metrics and Associated Delays

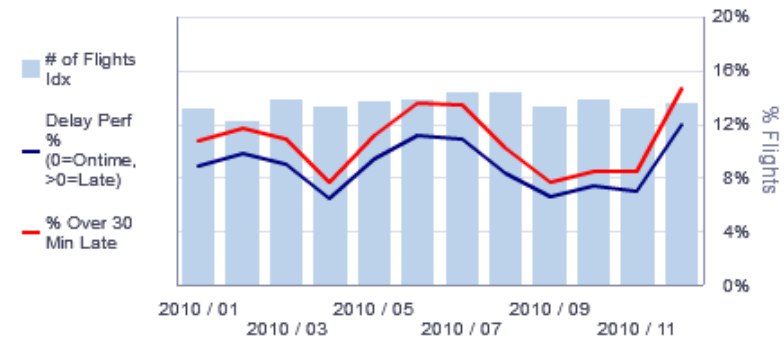
Orig Airport	# of Flights	% of Total	Passenger - Miles (M)	Delay Perf %
All Orig Airports	3,709,454	100.0%	328,034	9.1%
Midwest Region	691,998	18.7%	48,525	10.8%
East North Central Division	491,158	13.2%	33,998	10.9%
West North Central Division	200,840	5.4%	14,527	10.4%
Northeast Region	423,226	11.4%	44,389	8.2%
South Region	1,492,575	40.2%	107,799	9.9%
East South Central Division	177,198	4.8%	7,199	! 11.3%
South Atlantic Division	821,345	22.1%	67,067	9.8%
West South Central Division	494,032	13.3%	33,533	9.5%
West Region	1,101,655	29.7%	122,759	7.6%



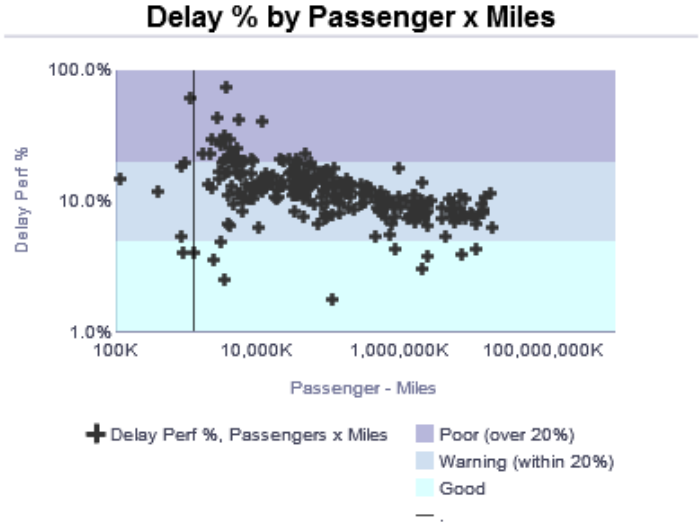
Late Flight Trends

By Time and Performance % Thresholds

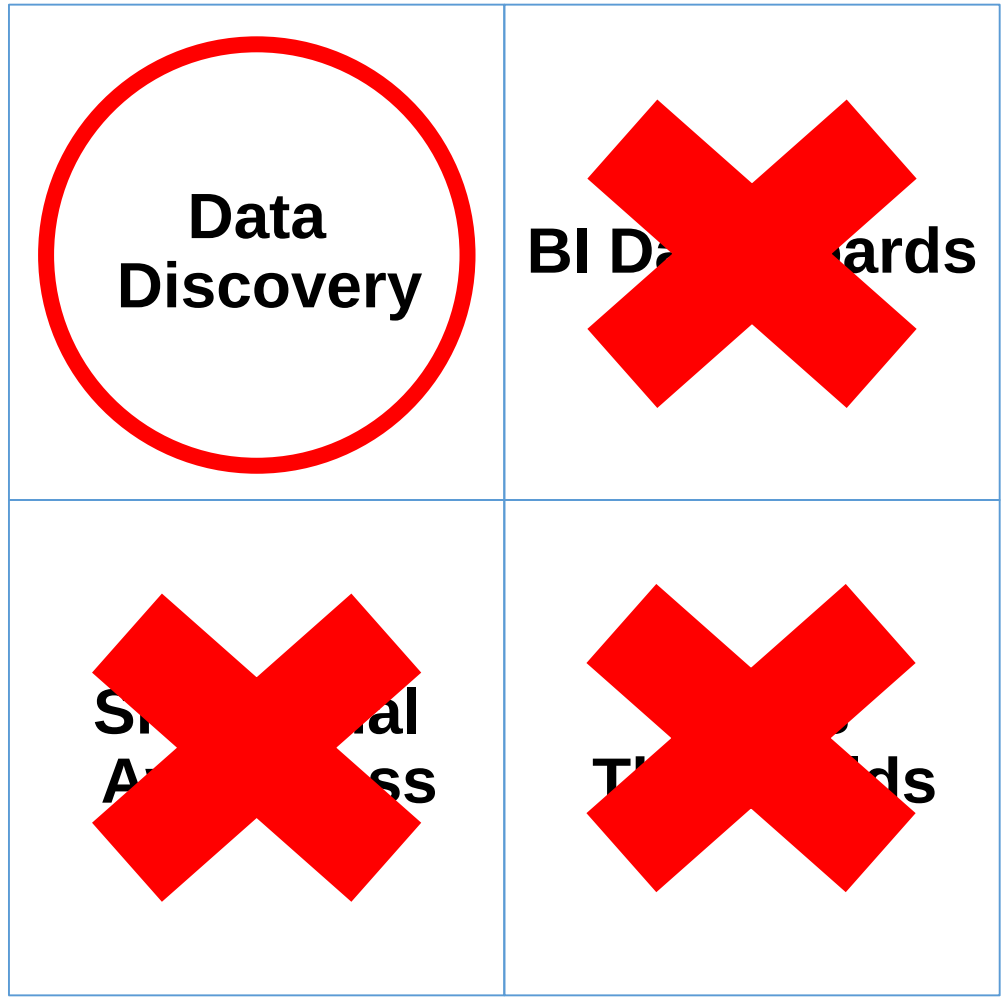
All Orig Airports



Orig Airport: All Orig Airports



Data Visualization Scenarios



Discovery - Explore vs Pioneer

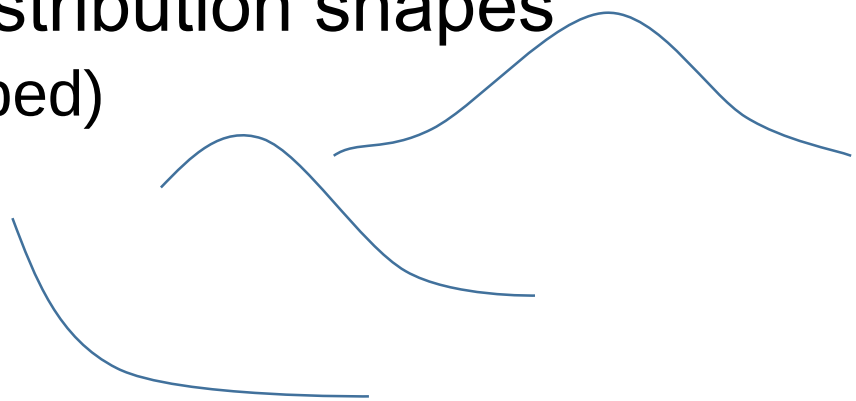
True Discovery

Data Discovery Sequence

- "Skim" the entire data set to get a sense of its size and scope
- "Read" the data set a **second** time more carefully
  - Identify facts/measures
  - Transaction/event records included?
  - Identify major dimensions
- Make a list of potentially important or interesting business issues/implications
- Compare your original business issues with your new list
- Apply useful frameworks
- Transform data and add new data
- Apply useful frameworks

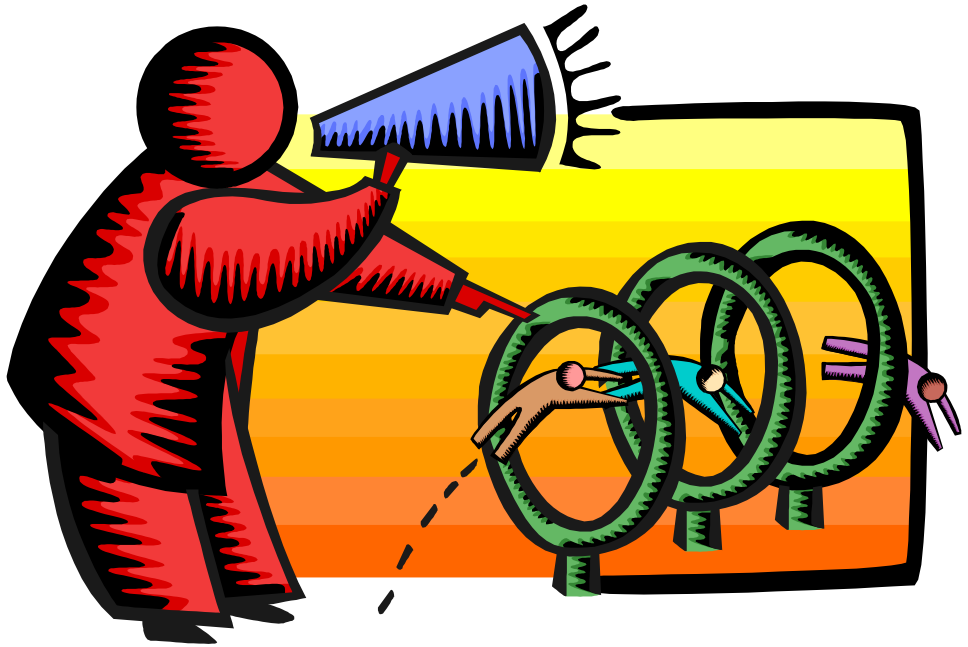
Understanding Measures for Exploration

- Aggregation method is important
- If use average, also add a bucketed measure
- Compute differences
- Understand data's natural distribution shapes
  - Normal distributions (bell shaped)
  - Log-normal distributions
  - Exponential distributions



- Average has strong meaning only for normal distributions
- Outlier identification & treatment are important for non-normal distributions

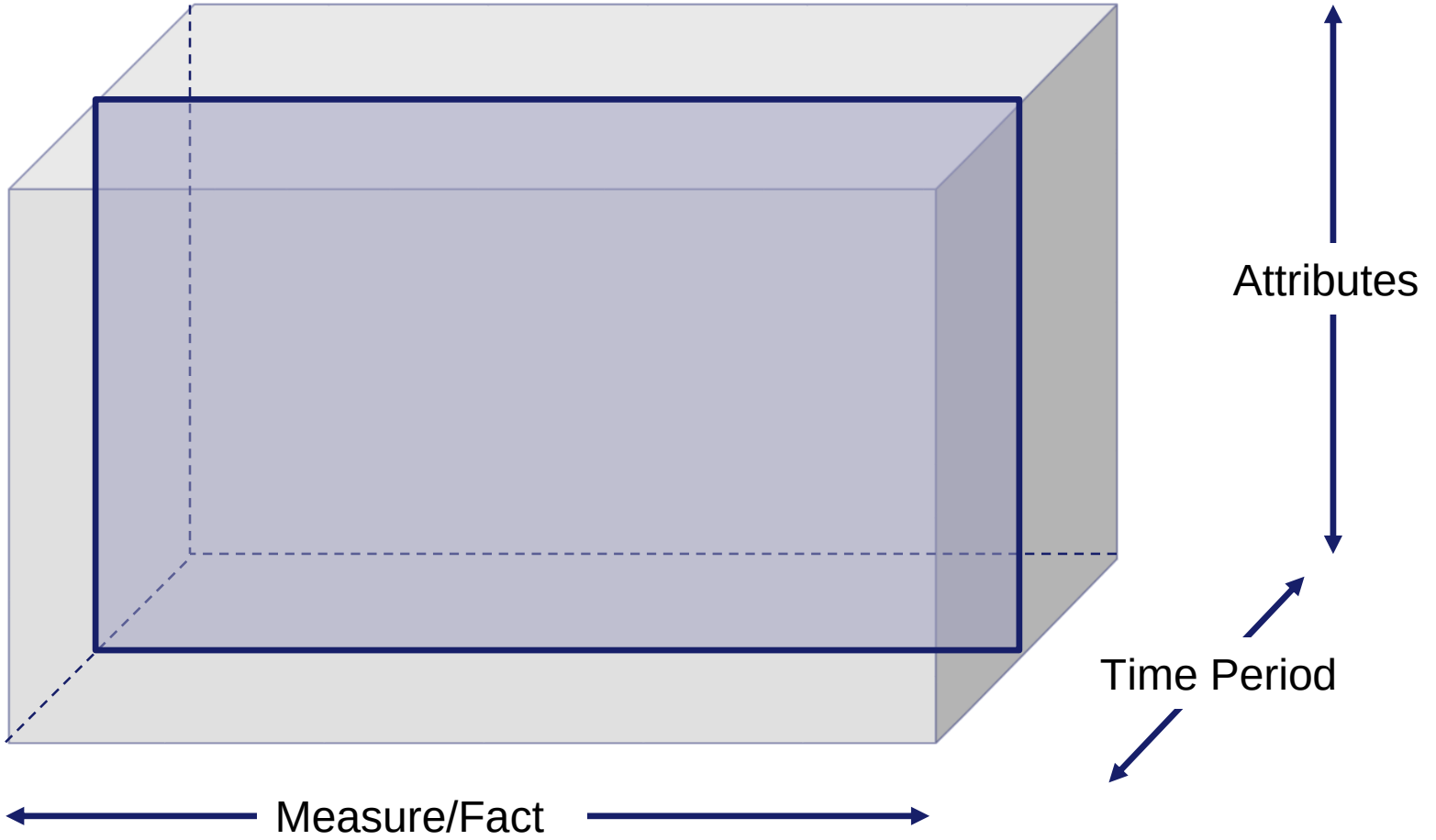
Demo



An Example Useful Framework

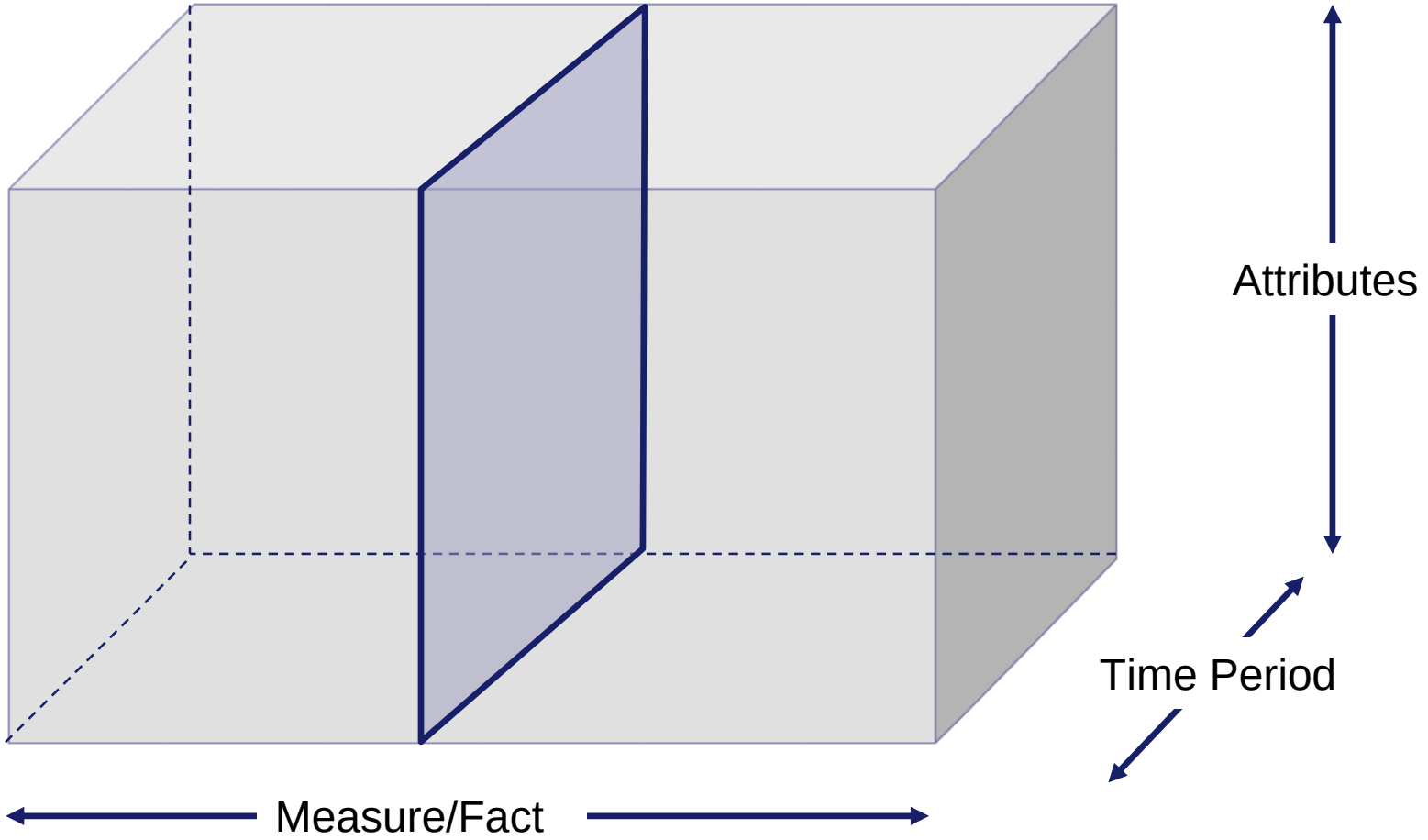
Position Analysis	Performance Analysis	Flow Analysis
static	period of time	period of time
descriptive	results	change in single asset/resource
relative/comparative	fixed vs. variable	sources and uses
balance sheet	P&L	cash flow
strength/weakness	bottom line/zero based	change over time
portrait	motion picture	narrative

Position Analysis



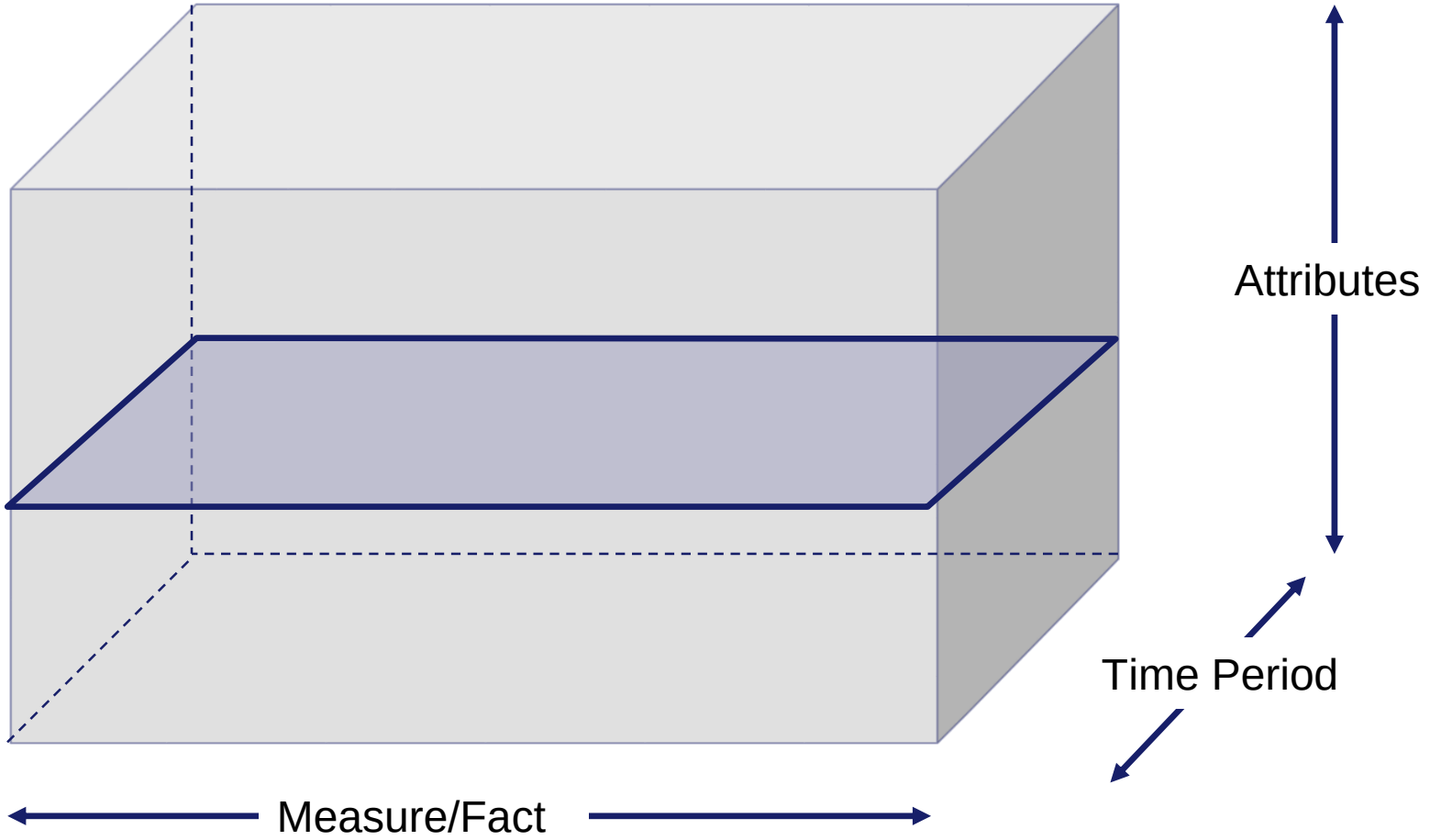
Bar Chart
Scatter Plot
Treemap

Performance Analysis



Bar Chart
Scatter Plot
Line Chart
Area Chart
Trellis

Flow Analysis



Line Chart
Area Chart
Trellis
Waterfall

Well Established Frameworks

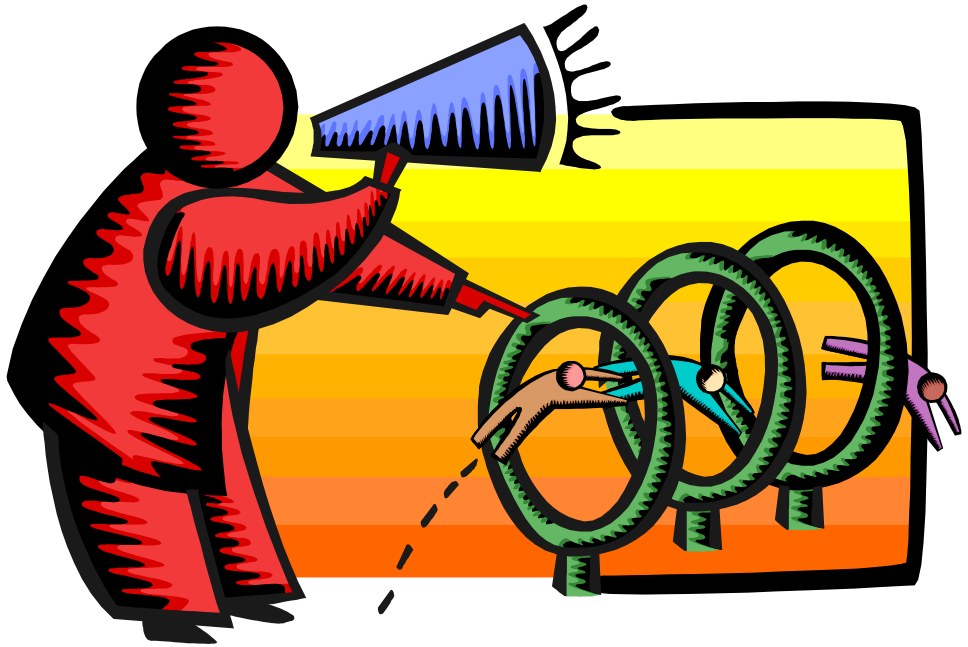
- Key Performance Indicator (KPI) Development (business)
- Root cause analysis (science)
- Diagnostic analytics (science)

Dimensional Columns

High number of factors / cardinality	**Lowest Grain**	**Trend/cycle** **Correlation** **Outlier**
Low number of factors / cardinality	**Trellis**	**Comparative** **Correlation**

Flat Shaped

Demo

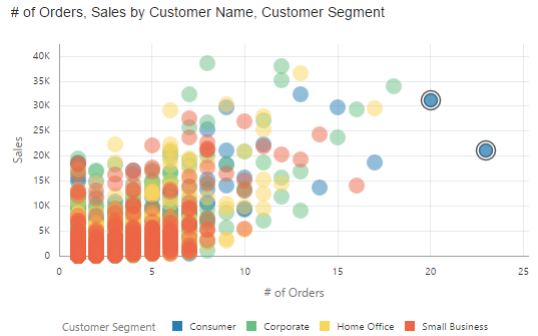


Keys to Data Discovery

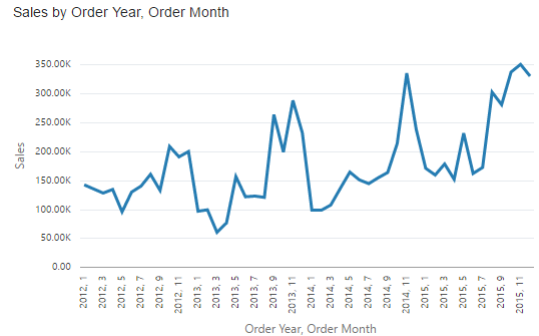
- Identify your main topic of interest with a performance tile
- Summary
- Evaluating a fact or a dimension?
  - Sales analysis
  - Customer or product analysis
- Fact analysis
  - Find lowest grain
  - Flat low distribution
  - Event or transaction
- Look for clustered distribution
    - Scatter with points as event in fact table
      - Set fact on X axis and response variable on Y axis

Major Types and Uses of Graphs

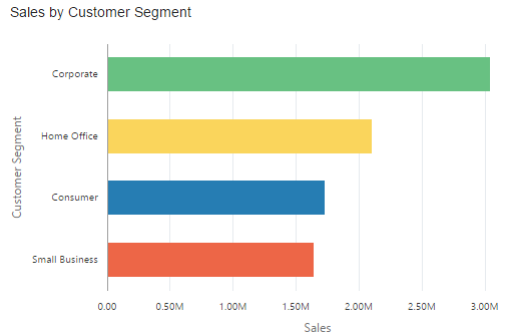
- Scatter plot – outlier detection



- Line graph – time based measures. Looking for trends and patterns



- Bar graph – comparison analysis



Starting with Data Discovery

- Begin either with a specific question or a framework
- Avoid “wandering around”
- Most of your visualizations will not produce new insights
- Move quickly through visualizations
- Be prepared to open a lot of browser tabs

Finding is not Explaining

- Process of interaction has a huge impact on the contextual understanding of an insight
- When someone discovers something, they believe it more
- Human Cognition Biases

Oracle Test Drive

- Free to try Oracle BICS, Oracle Advanced Analytics
- Go to www.vlamis.com/td
- Runs on Oracle Cloud
- Test Drives for:
  - Oracle BICS
  - Oracle Advanced Analytics (initially Oracle Data Mining)
- Once sign up, you can access for 24 hours
- Click by click script included, but can go “off road”
- Faster and easier than official Oracle “trial web account”

Using BICS Test Drive on Oracle Cloud

1. Go to www.vlamis.com/td

Select the cloud service you wish to try: *

- ○ Oracle BI Cloud Service - Visual Analyzer
- ○ Oracle Advanced Analytics - Oracle Data Mining

Your Name *

Email Address *

@

This must be your corporate email address. Requests from free mail services such as Gmail will be declined.

Company Name *

Daytime Phone Number *

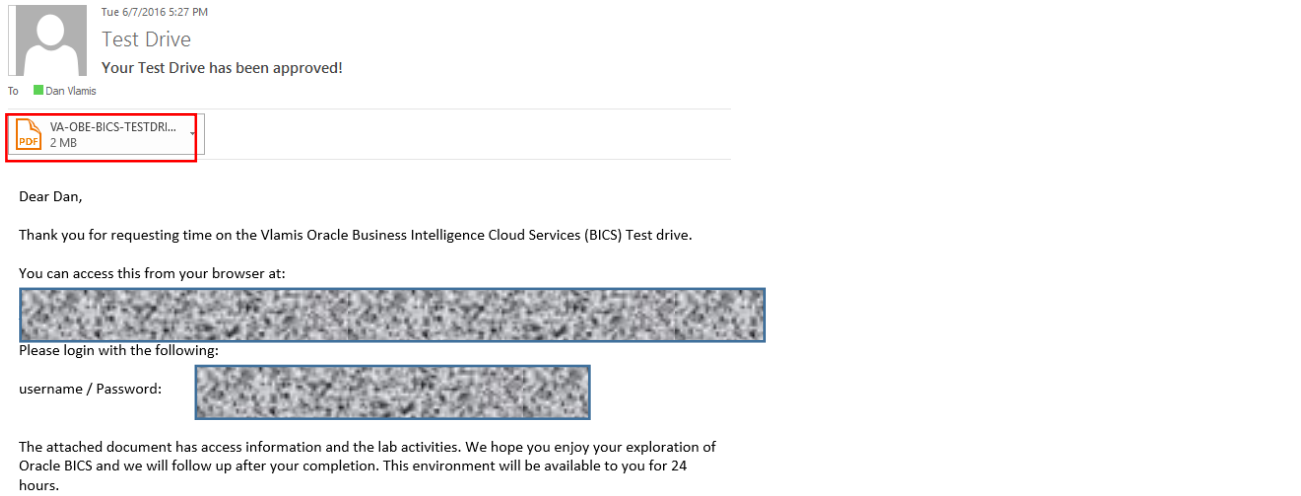
(###) ### ####

US phone number where you can be reached during normal business hours.

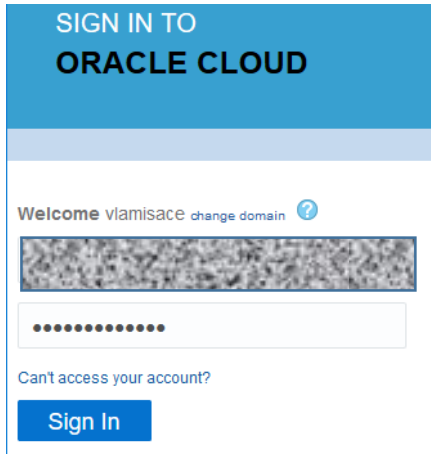
Comments

Submit

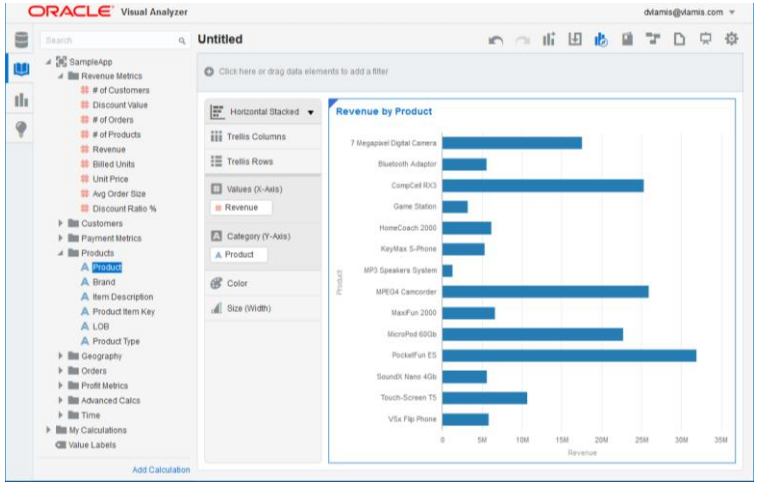
2. We send you email with userid/pw and script



3. Sign into BICS



4. Use PDF file script or go off-road!

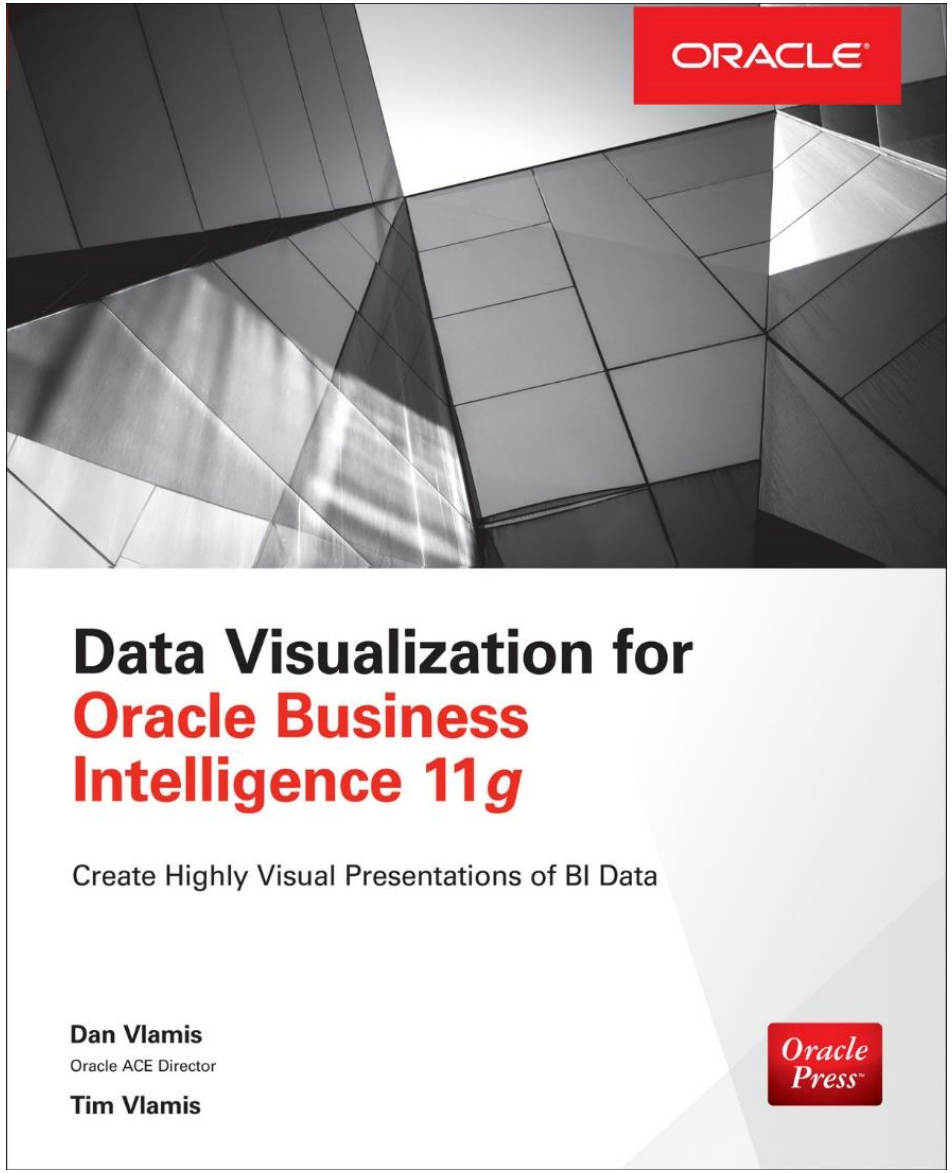


CELEBRATING EXCELLENCE vlamis 25 YEARS SOFTWARE SOLUTIONS

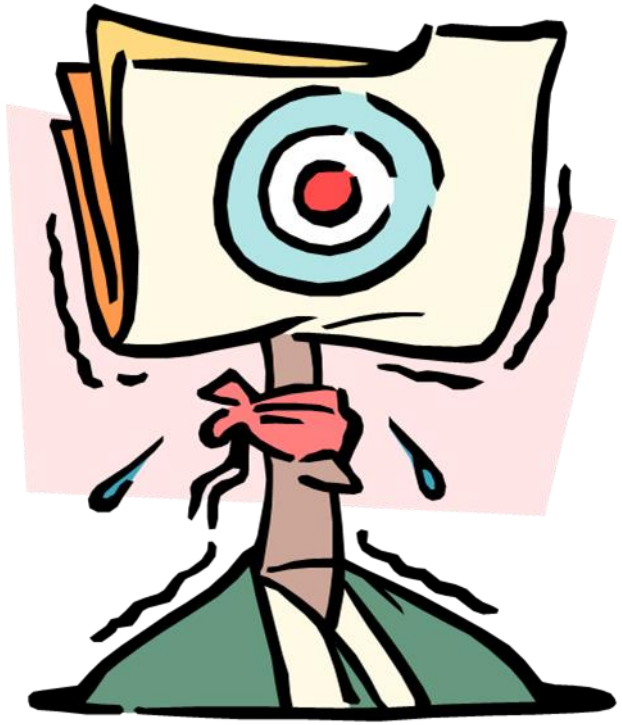
Drawing for Free Book

Add business card to basket

or fill out card



Questions?



Thank You!

Data Visualization Best Practices

Tim Vlamis
tvlamis@vlamis.com
www.vlamis.com
@timvlamis
@VlamisSoftware