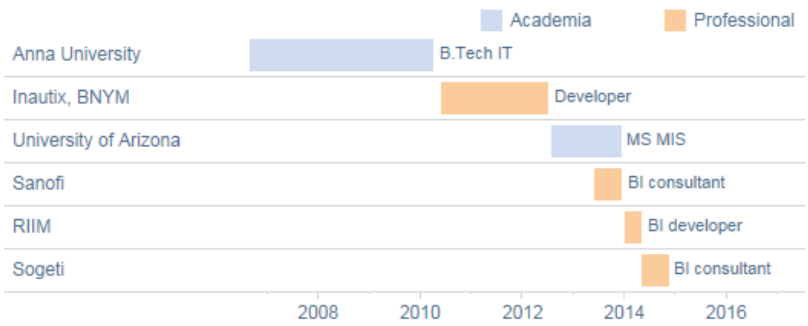
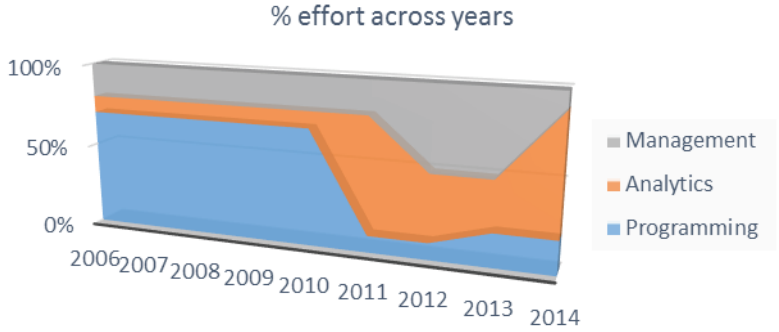
Arunkumar Navaneethan
arunkn79@gmail.com
520-369-1363 | 1255 E University Dr #341 AZ 85281

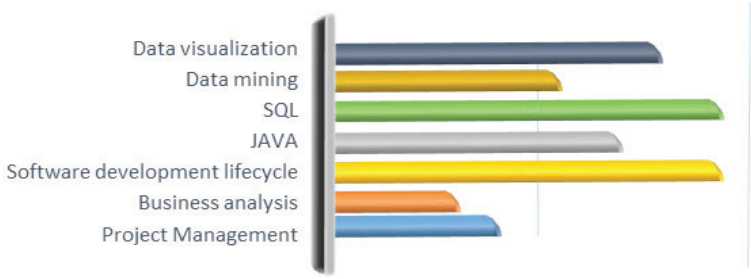
Scan QR code to visit
http://arunethan.com



- Experienced in Business Intelligence applications like Tableau, Microstrategy, OBIEE, Spotfire, IBM Cognos
- Worked in database systems like Oracle, DB2, MYSQL and SQL Server
- Worked in financial services, insurance and pharma research domains
- Strong JAVA web development background
- Practical understanding of data warehouse and OLAP concepts

Strong foundation in technology, analytics and good understanding of project management

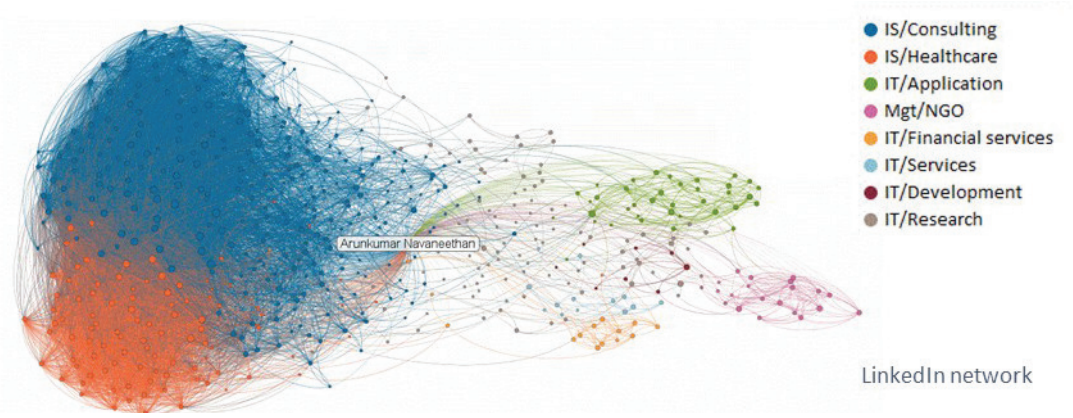




Expertise on a scale of 1, low to 10, high

Zeal to learn and excel in new technology by business requirement

Large professional network
Regular blogger on Business Intelligence and enterprise innovation specific topics at http://arunethan.com/



LinkedIn network

Tableau Interactive Dashboard Resume
www.arunethan.com/?page_id=101

Result driven technology professional with refined business skills and a successful track record in application development, database management and information systems management

- MS in Management Information Systems; University of Arizona; Eller College of Management; (12/2013)
- B.Tech in Information Technology; Anna University, Sri Sairam Engineering College; Chennai India; (4/2010)

Crestbrook, Nationwide Insurance (through Sogeti) **Consultant, Business Information management**
SQL Server, Report Management, Tableau, Master Data Management (May 2014 to Present)

- Rationalizing approaches to provide a single version of the truth for enterprise reporting
- Creating T-SQL, PL/SQL scripts to interface data between enterprise applications
- Developing reports in Tableau and Excel to support internal financial compliance and sales departments growth assessment
- Evaluating business intelligence applications to create standardized approach towards report development

American Realty Capital Properties (through RIIM) **Developer, Business Intelligence**
SQL Server, Tableau, Microsoft Cubes, Dimensional modeling (January 2014 to May 2014)

- Analyzed transactional and dimensional data present in SQL server and Microsoft CUBE using Tableau
- Created standard reports for property management and internal strategy group
- Enhanced the reporting environment performance by reviewing the Microsoft Cube and Tableau server configurations

Sanofi (through University of Arizona) **Consultant, Information System Management**
Innovation management Application, SharePoint (June 2013 to December 2013)

- Interacted with cross functional managers from USA and Europe to create a web application to manage innovation within the organization using Spigit application platform
- Explored the functionalities of in-house data management and business intelligence applications, capitalized the out of the box features to match the business needs
- Pitched a proof of concept using technologies like SharePoint, Yammer, Denodo data virtualization and in-house text analytics platform to manage internal events activities and engagements

University of Arizona **Graduate student consultant, Information system management**
MySQL, Oracle, Data mining, Social media analytics, Data visualization (August 2013 to December 2013)

- Reviewed and enhanced the database system of the marketing department, updating the stored procedures, triggers, views of the Oracle DB system
- Performed predictive analysis using classification, clustering, association rule mining and network analysis; descriptive analysis using MySQL for back-end and Google visualization API for front-end
- Analyzed user behavior, demographics, device access patterns to enhance new acquisition and current members ease of web experience for StartupTucson.org venture company

Inautix, Bank of New York Mellon **Developer, Application Development**
Mainframe, J2EE, DB2, Oracle DB, IBM Lombardi BPM (July 2010 to July 2012)

- Involved actively throughout the software development life cycle; planning, requirement gathering end-to-end support, individually managed the application documents, recorded project progress
- Created reporting using Fusion chart API and parsed data from Oracle DB to JSON format using Java programs
- Handled change request, bug fixes, testing and test region code deployments. Lead the production support activities during the vendor transition and the internal group training sessions
- Developed new workflows, migrated and redesigned code from Java to Lombardi authoring environment
- Interacted with stakeholders like client, business analyst, business process management architect and database architect

Project Samples

Social Media Analytics

	Mashable	engadget
Type of articles	Technology World News Business Lifestyle Entertainment	Technology
Twitter Ranking	327th	2,162nd
Social media comparison	Twitter : 3.45 M Google+ : 2.80 M Facebook : 1.30 M	Twitter : 725,845 Google+ : 1,703,977 Facebook : 268,596

Balanced Scorecard

Content Curation via Twitter

Financial

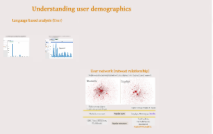
Improve profitability

Customer

Deliver customized articles

Internal Process

Improve social media presence

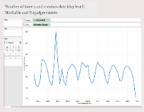


Learning & Growth

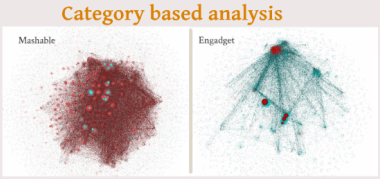
Increase employee involvement

Articles trends and performance

Article's temporal analysis



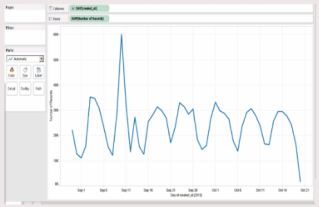
Category based analysis



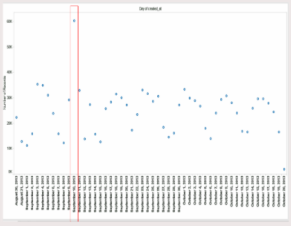
3.42	*Average degree*	3.17
6.31	*Average weighted degree*	11.21
25.6	*% of hashtag*	33.2
NYCC, iphone, comiccasts, toy, apple	*Trending topics (Eigenvector centrality based)*	smartnews, iphone5s, endagetiphone, iphone5c, iphone4s, iphone, apple

Article's temporal analysis

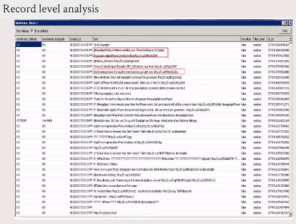
Number of tweets and creation date (day level)
Mashable and Engadget tweets



Number of tweets and creation date (day level)
Mashable and Engadget tweets

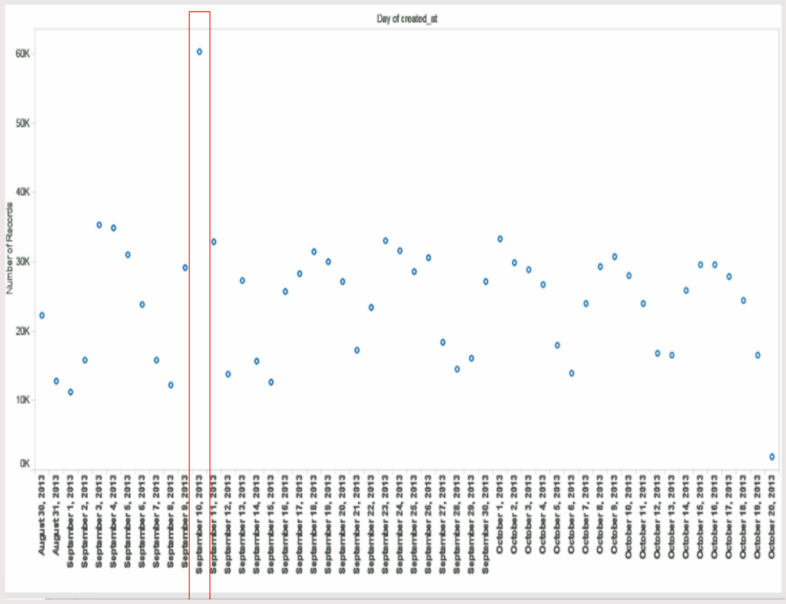


Record level analysis

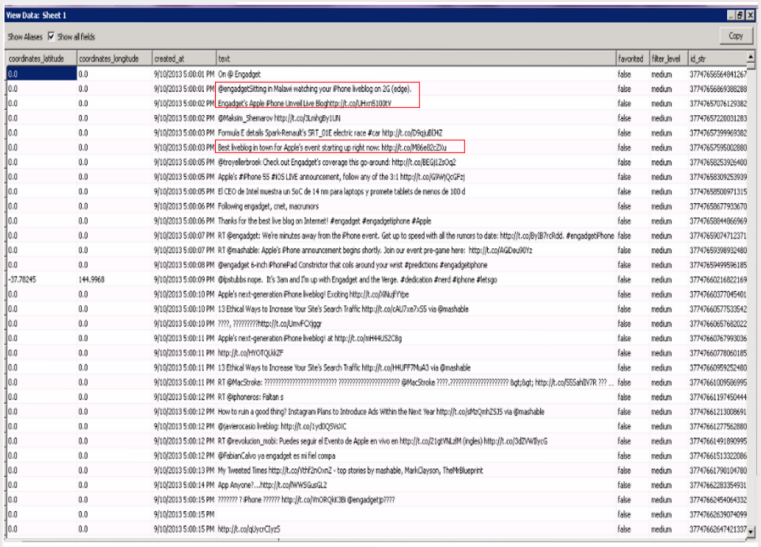


Engadget's Apple iPhone Unveil Live Blog
Best liveblog in town for Apple's event starting up right now
Apple's #iPhone5S #ios LIVE announcement, follow any of the 3:1

Number of tweets and creation date (day level)
Mashable and Engadget tweets



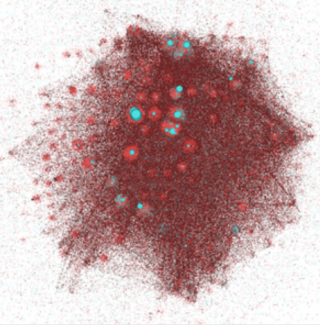
Record level analysis



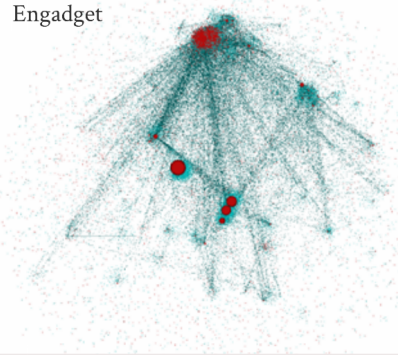
Engadget's Apple iPhone Unveil Live Blog
Best liveblog in town for Apple's event starting up right now
Apple's #iPhone5S #ios LIVE announcement, follow any of the 3:1

Category based analysis

Mashable

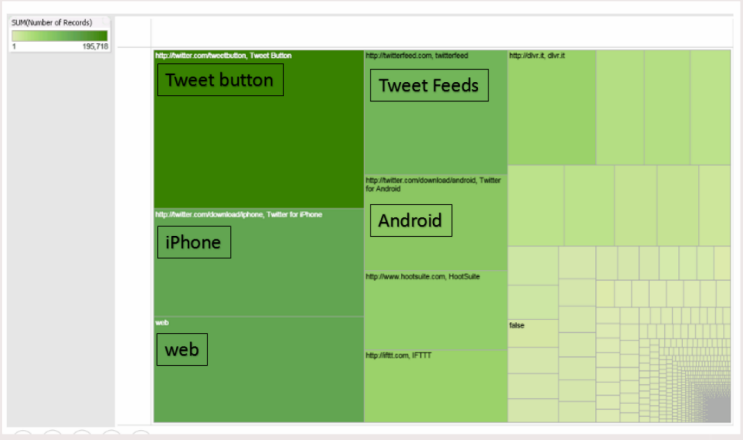


Engadget



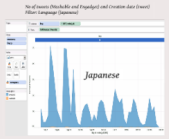
Mashable		Engadget
3.42	*Average degree*	3.17
6.31	*Average weighted degree*	11.21
25.6	*% of hashtag*	33.2
NYCC, iphone, comiccasts, toy, apple	***Trending topics*** ***(Eigenvector centrality based)***	smartnews, iphone5s, endagetiphone, iphone5c, iphone4s, iphone, apple

Heatmap of sources used to tweet
Mashable and Engadget tweets



Understanding user demographics

Language based analysis (User)



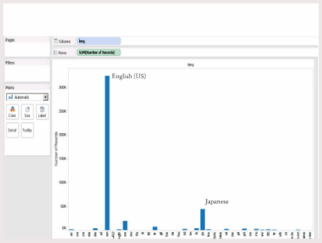
User network (retweet relationship)

Single mode directed weighted network; size: degree; Color: degree; Layout: openord

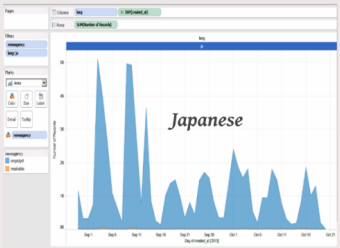
Higher average degree Lower average path length		Higher average weighted degree
Mashable, Comiccast	*Popular users*	Engadget, Morenograu, hatebu
CMC_Tech, HITECNews, iTechNews21	*Popular retweeters*	TechFeedLive, AusTechTweeter, dmxplosive1 MadeOfWWEAndHTC

Language based analysis (User)

No of tweets (Mashable and Engadget) and Languages (User)



No of tweets (Mashable and Engadget) and Creation date (tweet)
Filter: Language (Japanese)



User network (retweet relationship)

Single mode directed weighted network; size: degree; Color: degree; Layout: openord

Mashable		Engadget
Higher average degree Lower average path length		Higher average weighted degree
Mashable, Comiccast	***Popular users***	Engadget, Morenograu, **hatebu**
CMC_Tech, HITECNews, iTechNews21	***Popular retweeters***	TechFeedLive, AusTechTweeter, dmxplosive1 MadeOfWWEAndHTC


Arunkumar Navaneethan
Masters in MIS, Eller College of Management
University of Arizona
Innovation Management Projects
1. Innovation mapping initiative
2. INNOV8; ID8 innovation portal
Sanofi Summer Internship 2013
THE UNIVERSITY OF ARIZONA
SANOFI
Prezi


Innovation Management Projects
1. Innovation mapping initiative
2. INNOV8; ID8 innovation portal
Prezi

My Contribution

- Interacted with R&D Hubs, Genzyme, communications, Pasteur and other units leadership

Deliverables
- technical analysis
- Mock up dashboard

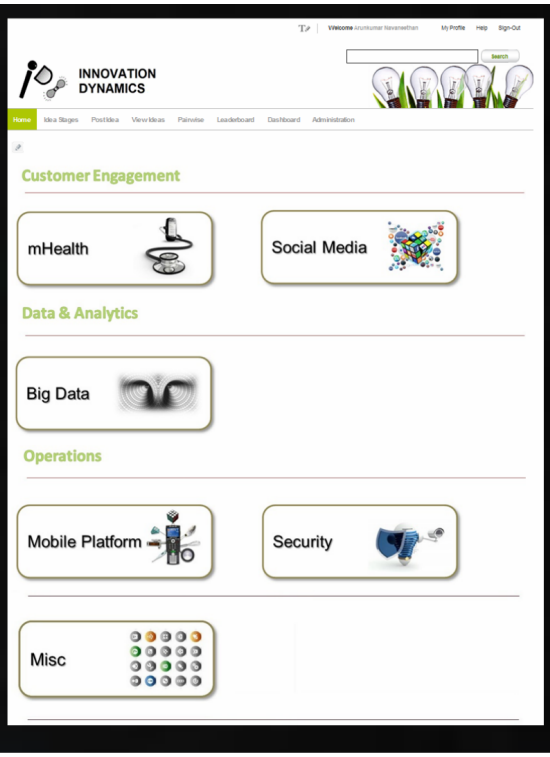
My Contribution

Created and designed the platform

Identified and managed platform issues

spigit

Prezi

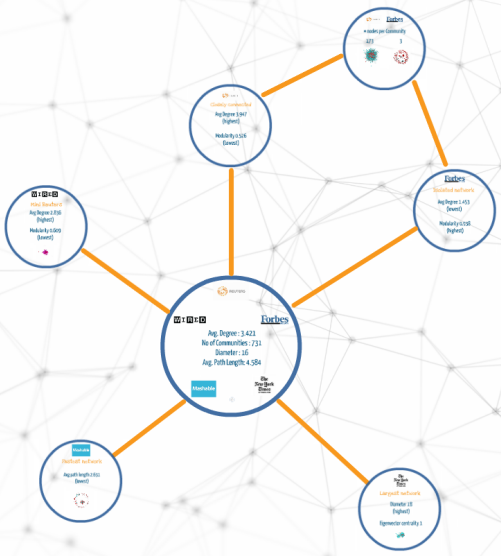


Prezi

User - Articles network analysis

Team Intelsprinters

Intel Sprinters



REUTERS

WIRED

Forbes.com

Avg. Degree : 3.421

No of Communities : 731

Diameter : 16

Avg. Path Length: 4.584

Mashable

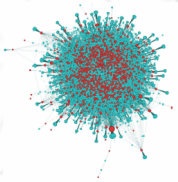
The New York Times NYTIMES.COM

REUTERS

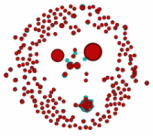
Forbes.com

nodes per Community

173



3

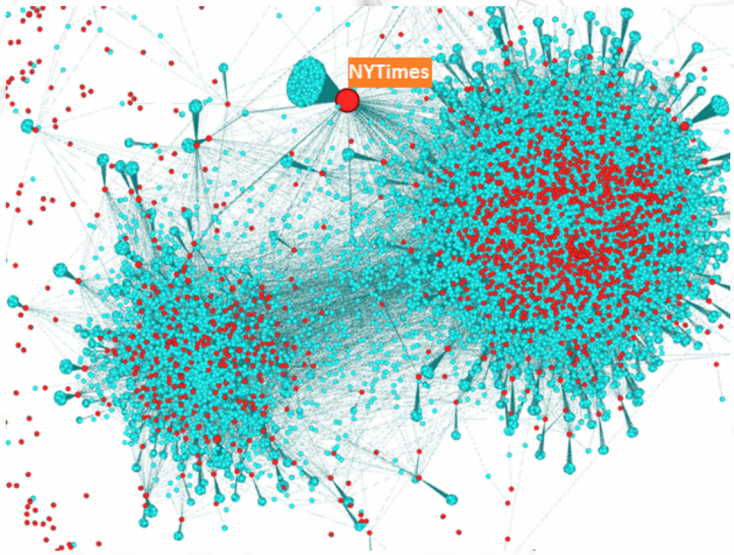


The New York Times
NYTIMES.COM

Largest network

Diameter 18
(highest)

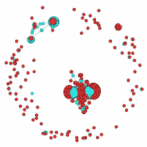
Eigenvector centrality 1

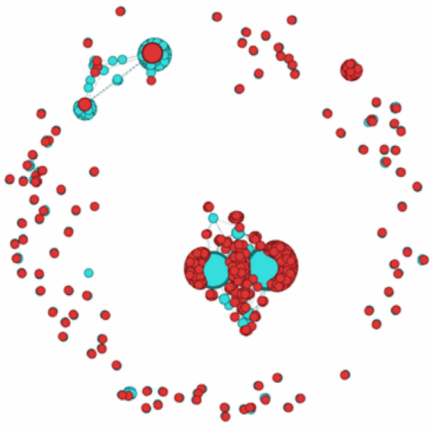


Mashable

Fastest network

Avg path length 2.631
(lowest)





WIRED

Mini Reuters

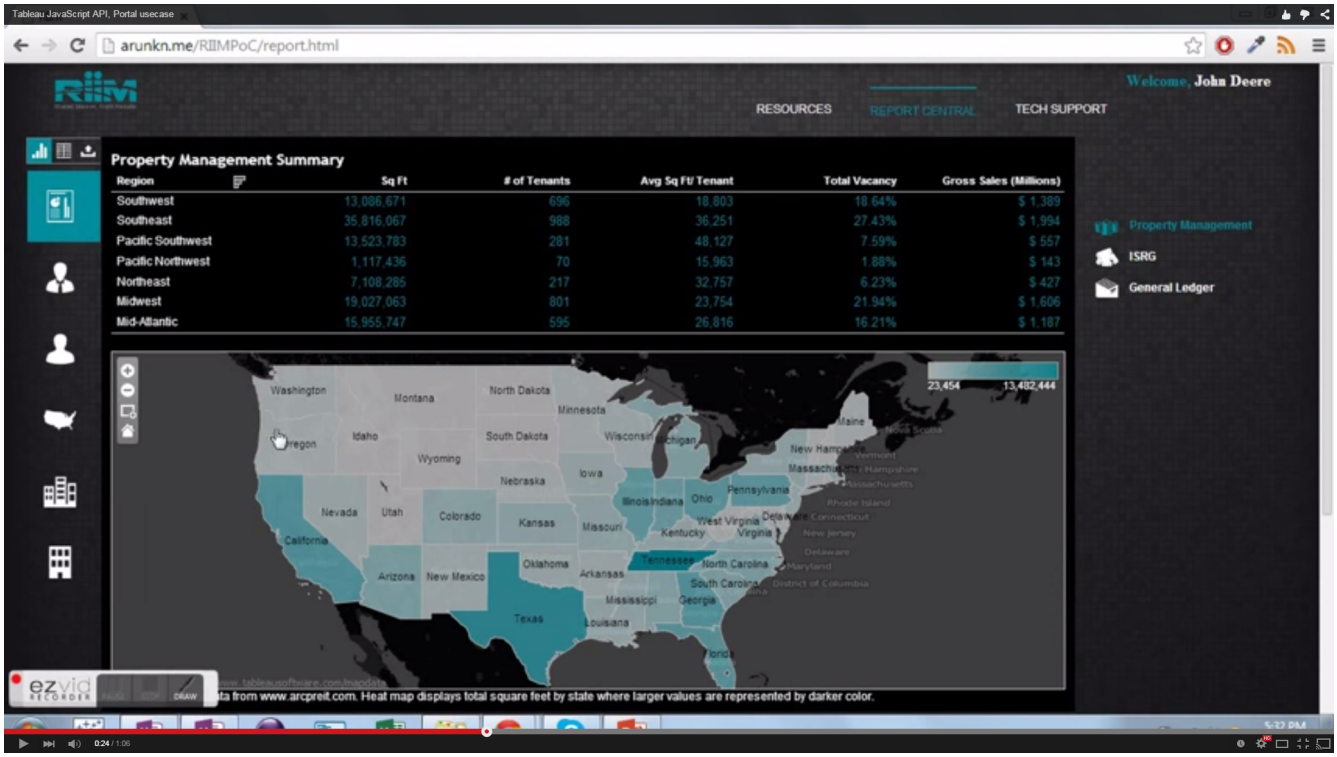
Avg Degree 2.836
(highest)

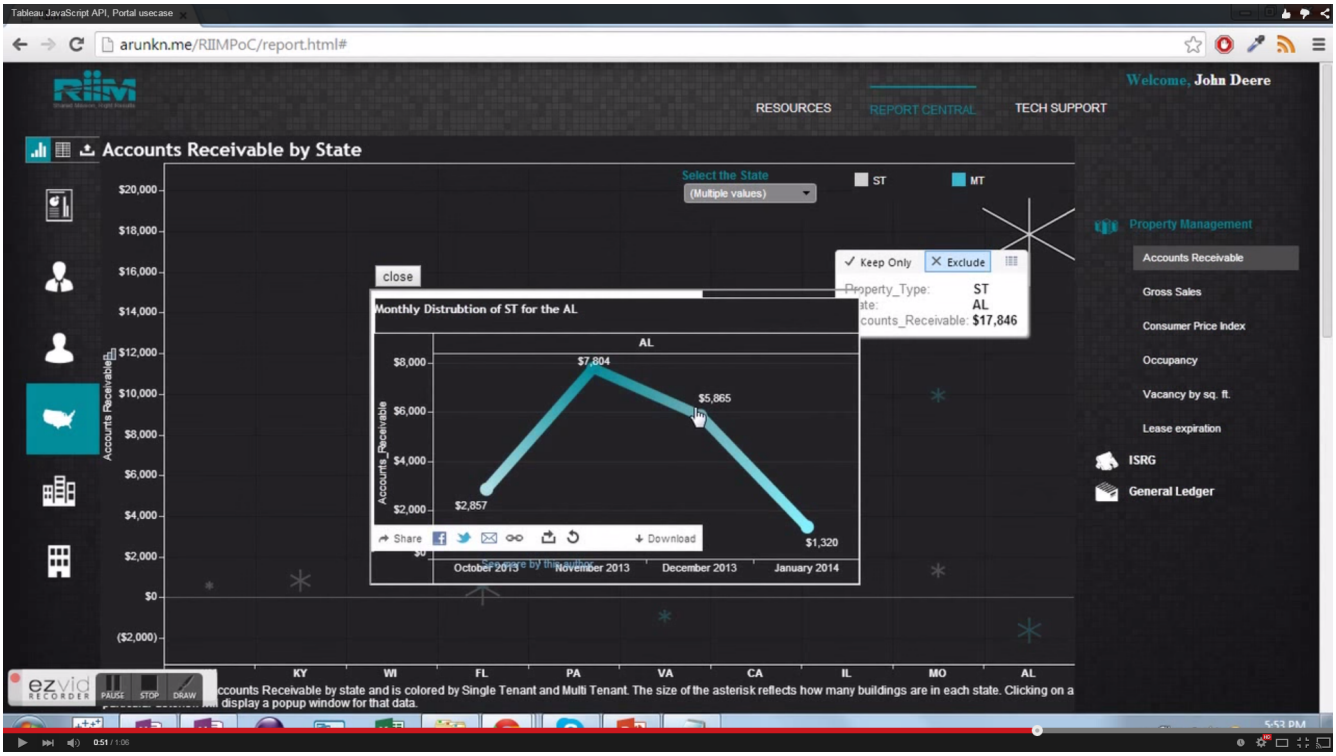
Modularity 0.609
(Lowest)

Tableau Experience Samples

Youtube demo of Tableau JavaScript API

https://www.youtube.com/watch?v=_91cpew085Q





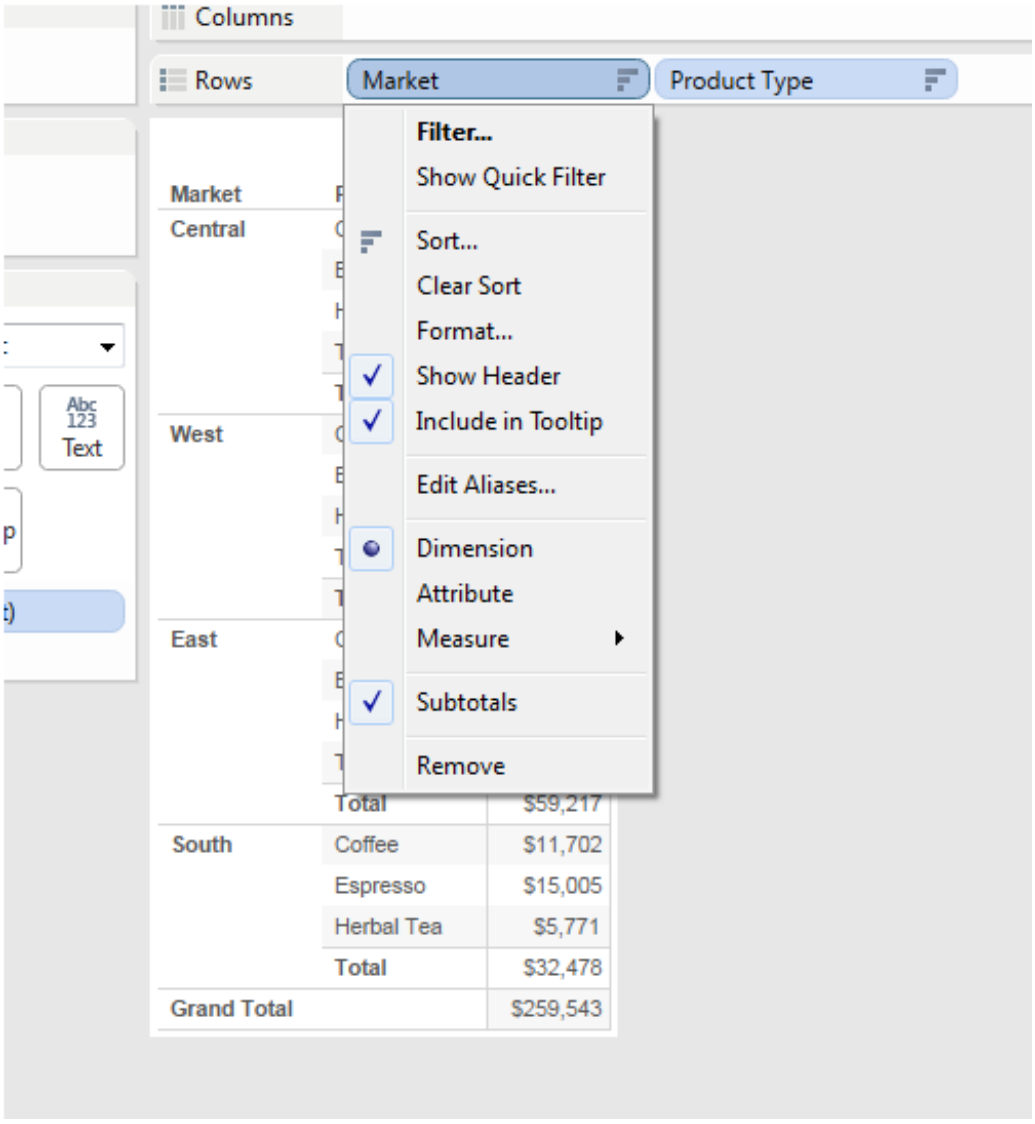
Sorting in Tableau

By Arunkumar Navaneethan

Basic sort

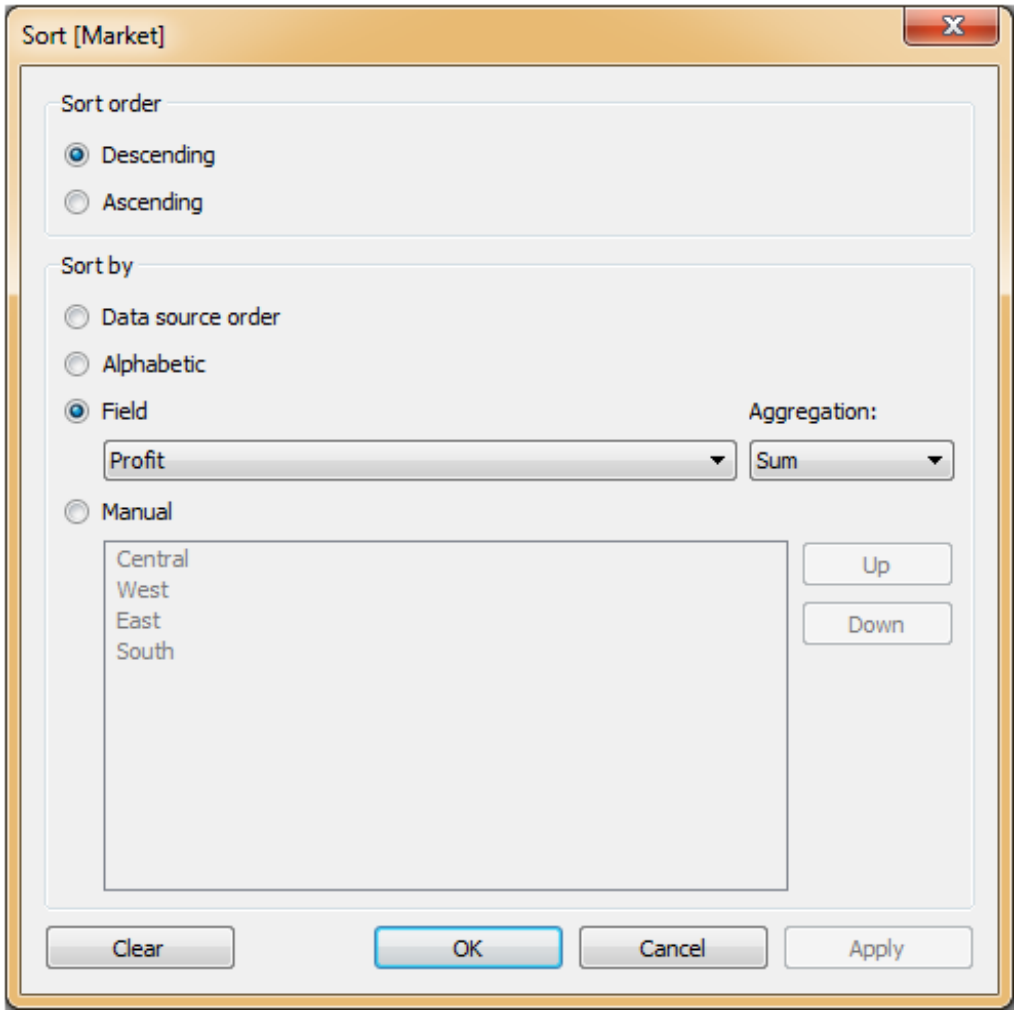
Right click on the discrete[1] field in the rows or columns shelf

[1] Sort can be performed only on discrete fields or measures which is converted to discrete



The discrete field can be sorting by

1. Data source (if the field is pulled from a SQL data, this order is based on sort orders applied to this relation)
2. Alphabetically
3. Based on another Field across different aggregation level

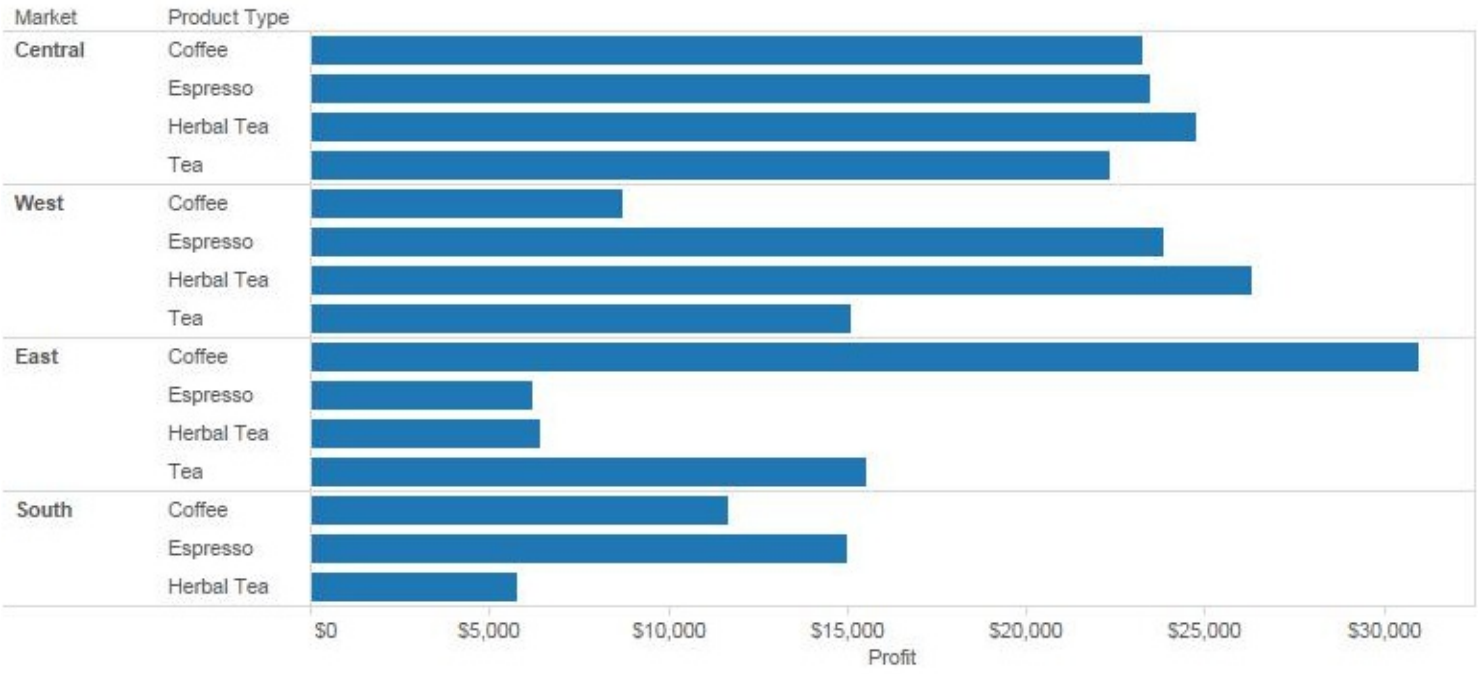


Non-Nested Sort

The nature of the sort function in Tableau is that sorting is done at the content level.

When a sort is done based on the SUM(profits) descending, you can notice that CENTRAL market is at the top. This because the aggregate SUM of CENTRAL is the maximum. Similarly COFFEE product type as a MAX of SUM(PROFIT) arranged descending. Hence it is at the top across all the Markets. The results are not sorted based on the Markets but to the entire context.

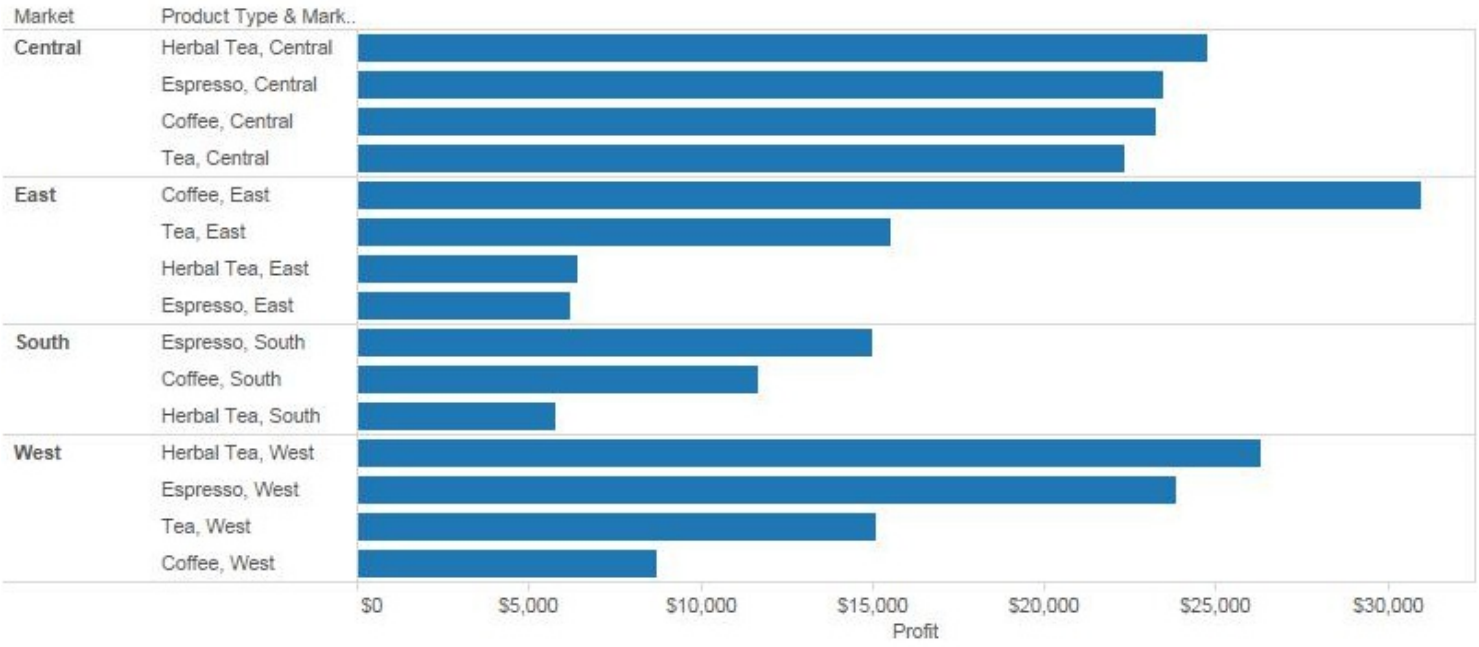
This is the default behavior within Tableau.



Nested Sort

If the results is expected to be sorted based on the markets, we have to create a nested sort using a work around.

In the dimensions pane, press the CTRL key, and then select the dimensions. Right-click on e of the selected fields and click on Combine Fields option. The newly combined field, Market & Product Type is created. Drag this field to the Rows panel and set the sort options. We will get a nested sort output.



Selected Sort

With the previous sort you can add more measures but the results will still be still sorted by PROFIT

Market	Product Type & Mar..	Profit	Sales
Central	Herbal Tea, Central	$24,757	$67,888
	Espresso, Central	$23,501	$59,701
	Coffee, Central	$23,264	$69,080
	Tea, Central	$22,330	$68,376
East	Coffee, East	$30,992	$56,641
	Tea, East	$15,558	$32,177
	Herbal Tea, East	$6,423	$41,361
	Espresso, East	$6,244	$48,397
South	Espresso, South	$15,005	$44,990
	Coffee, South	$11,702	$33,259
	Herbal Tea, South	$5,771	$25,677
West	Herbal Tea, West	$26,303	$72,288
	Espresso, West	$23,870	$69,908
	Tea, West	$15,098	$72,220
	Coffee, West	$8,725	$57,848

Tableau combination chart

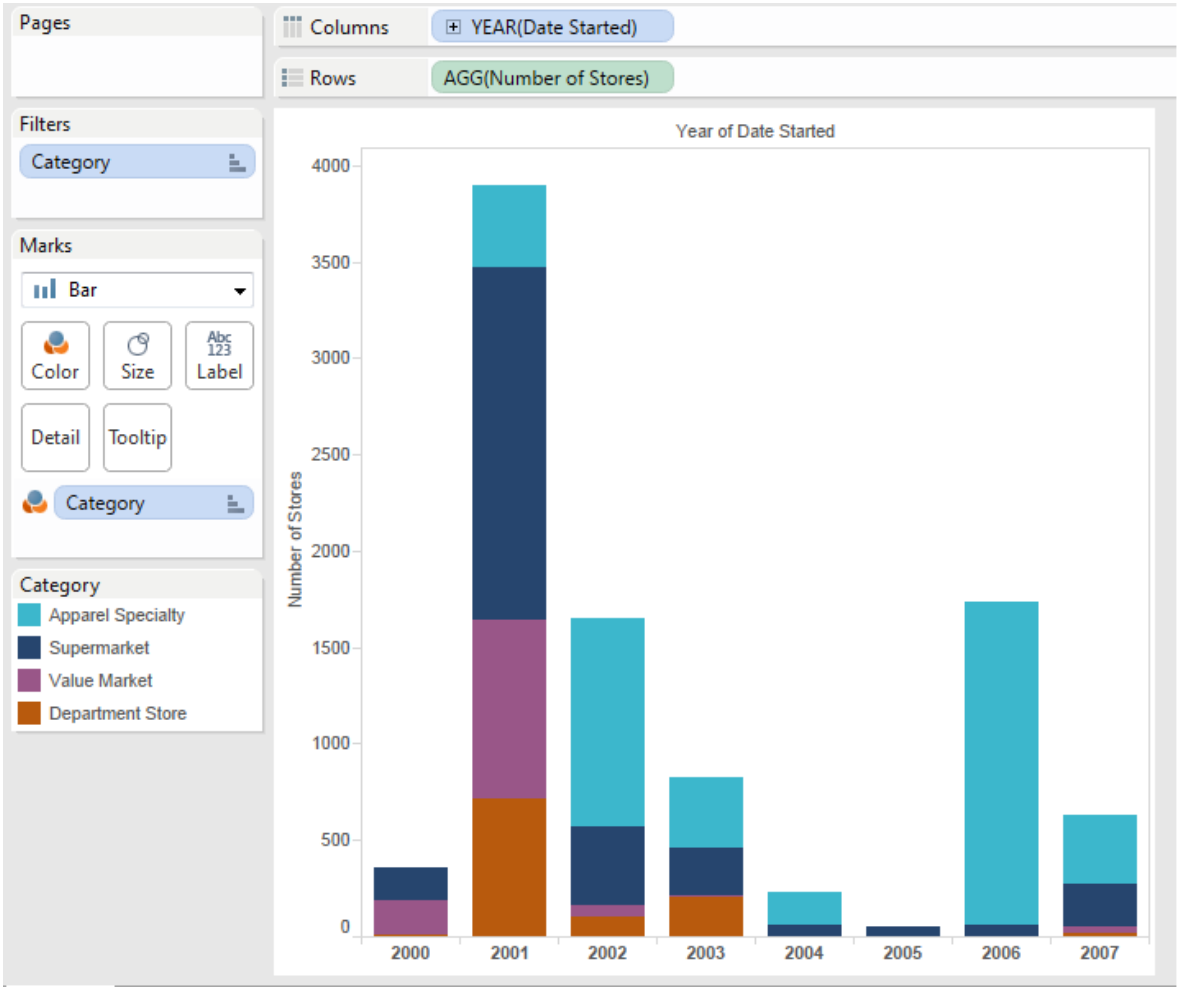
By Arunkumar Navaneethan

My Use-case

To display a chart with Number of stores across years, differentiating each year by category of the store. In addition, have an overlay that would show the total for the given year based on user selection. This should not modify the underlying chart.

Step 1:

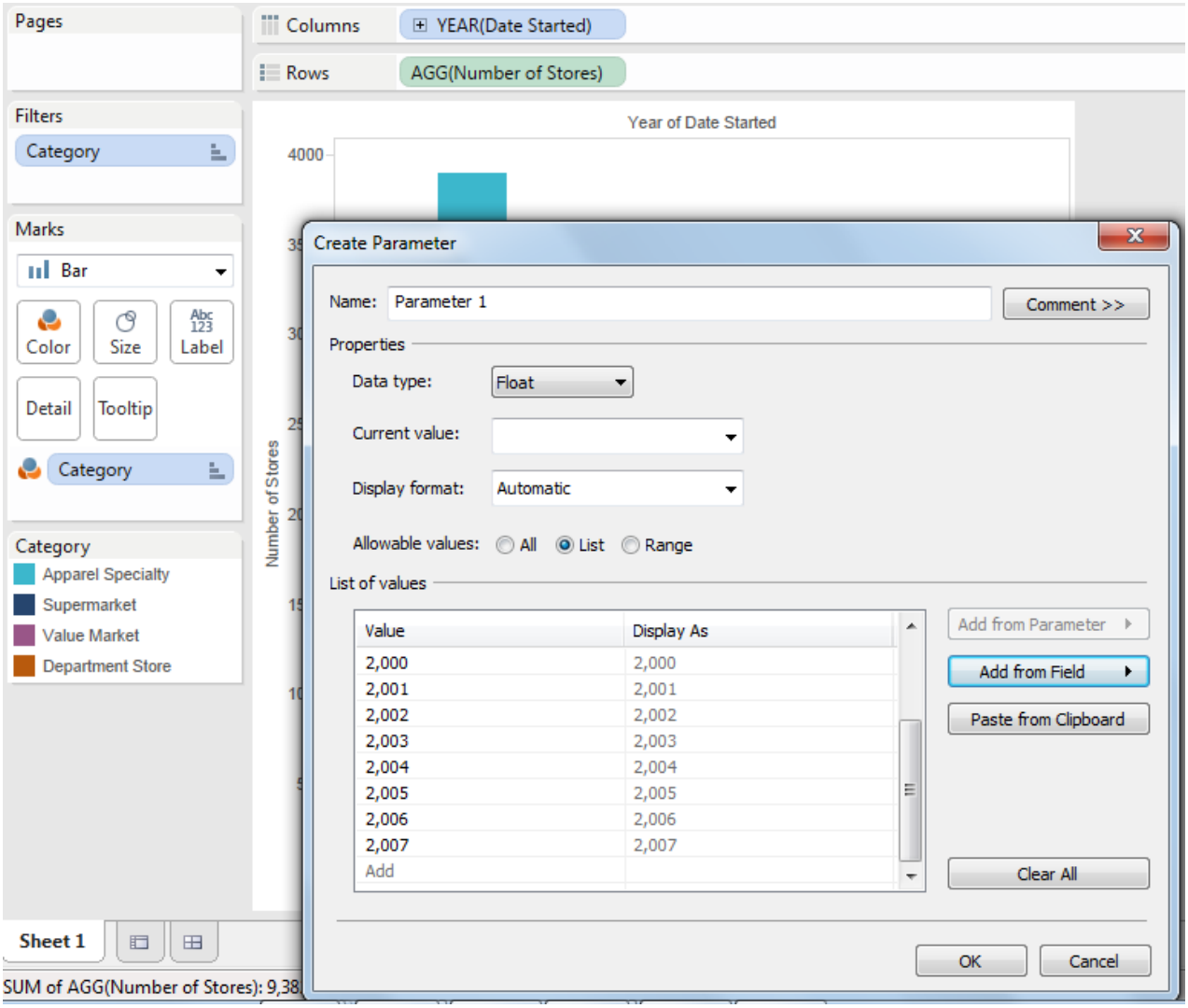
1. Create a Stacked chart



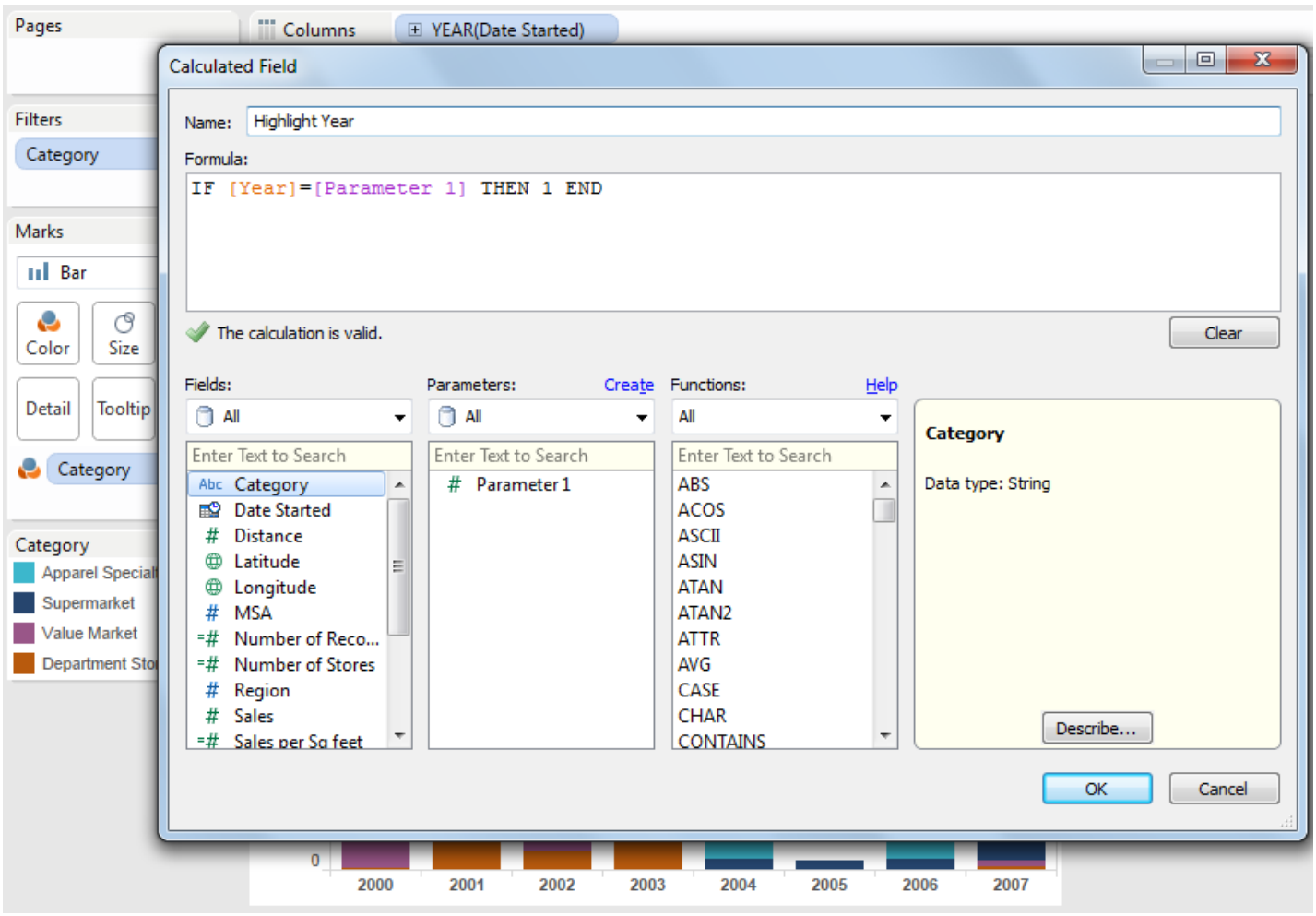
Rows: *Aggregate of number of stores*
Columns: *Years*
Filter: *The category of interest (Apparel specialty, supermarket, value market, department store)*
Color: *Category*

2. *Overlaying chart*

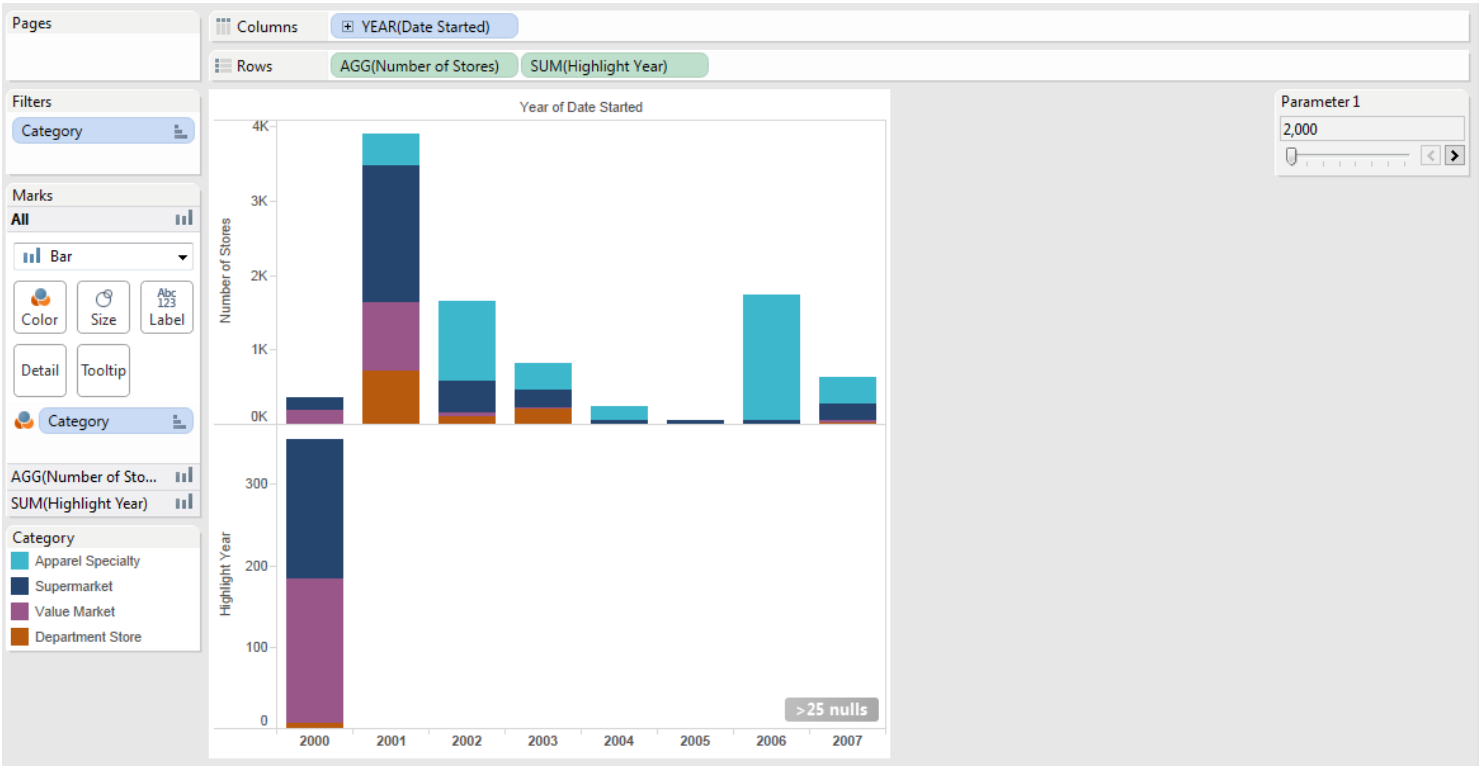
2.1. Create a parameter using Year field and select years from 2000 to 2007



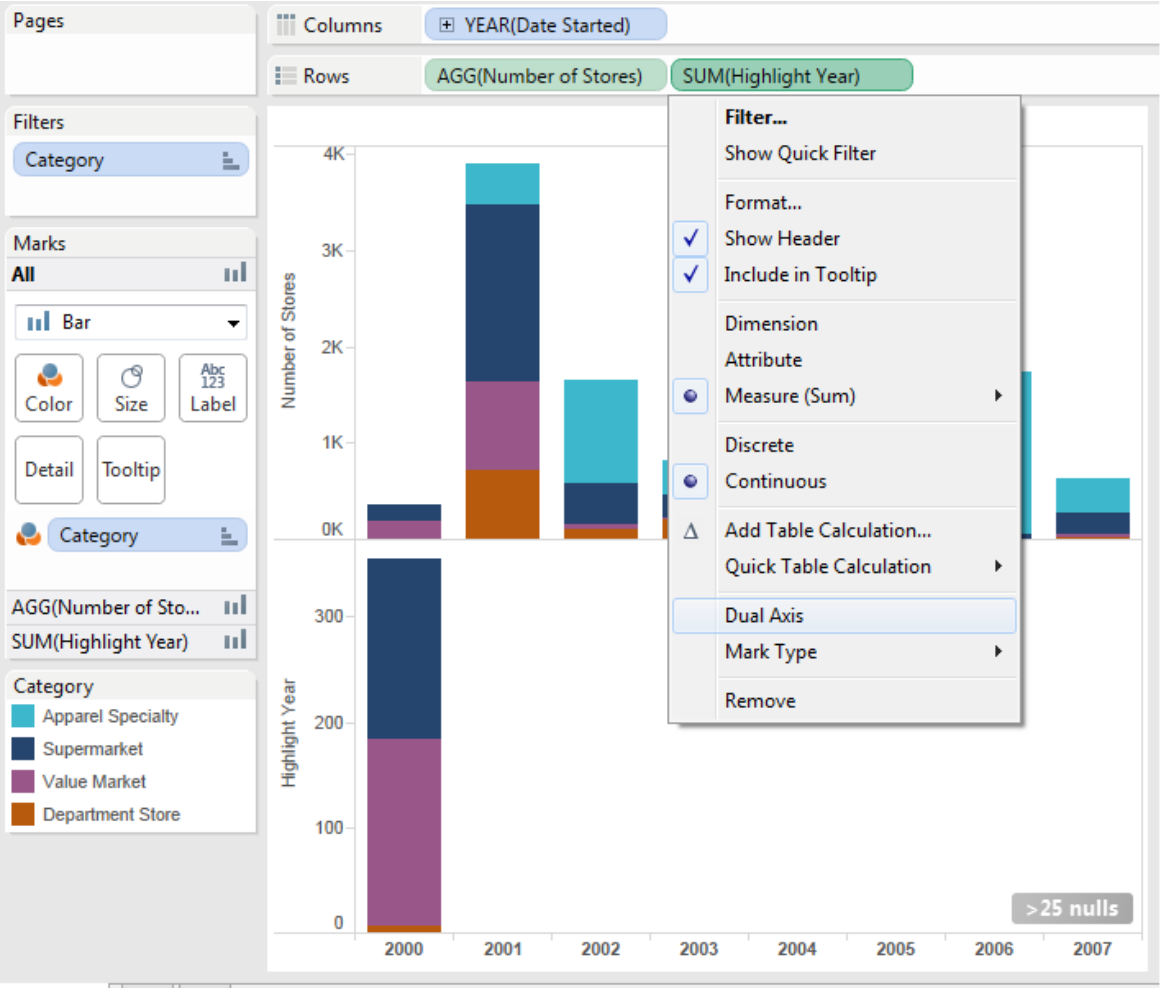
2.2. Create a calculated field Highlight Year (IF [Year]=[Parameter 1] THEN 1 END)



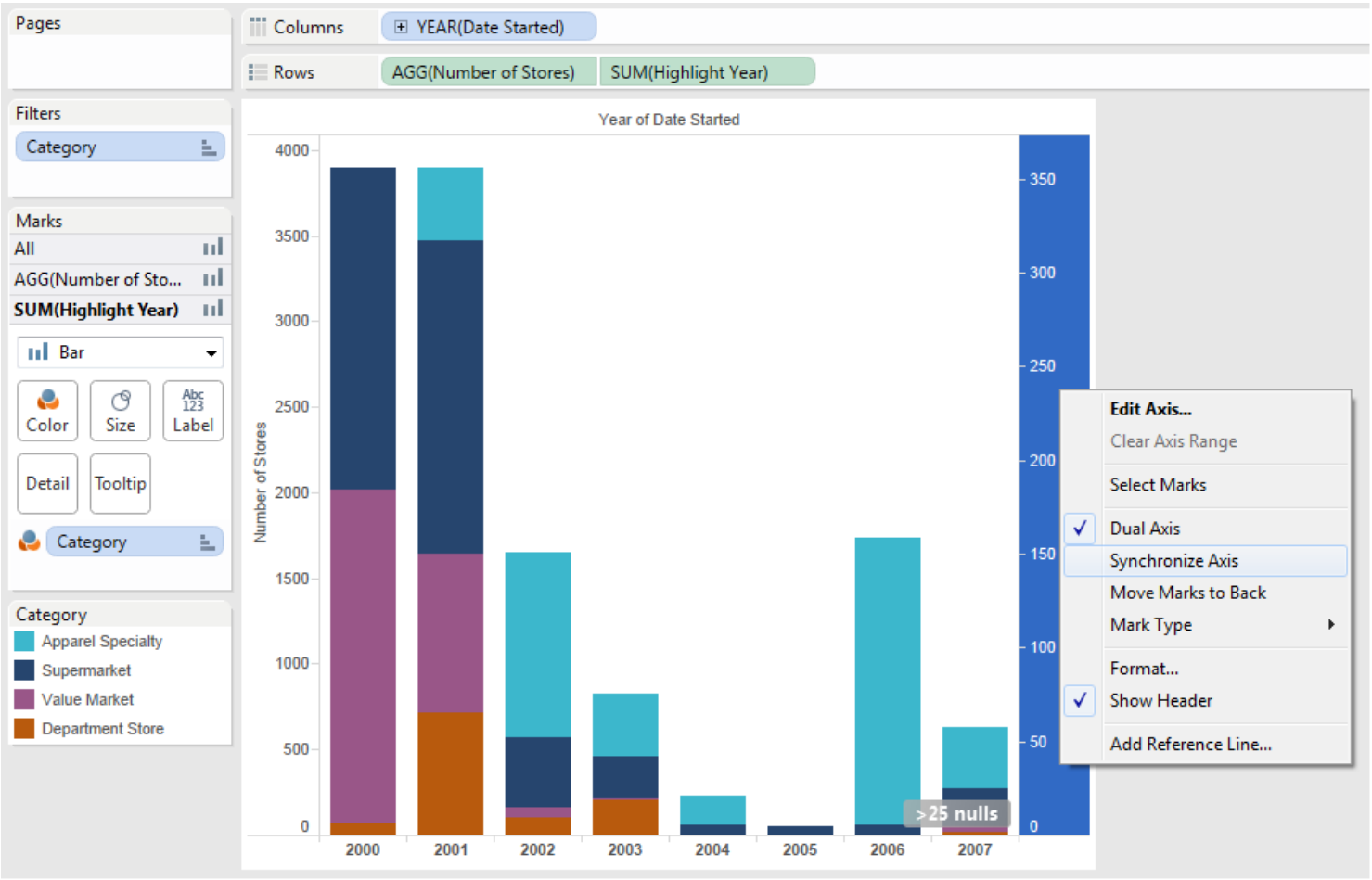
2.3. Display the parameter in the sheet and make it a slider, move the Highlight Year to rows shelf



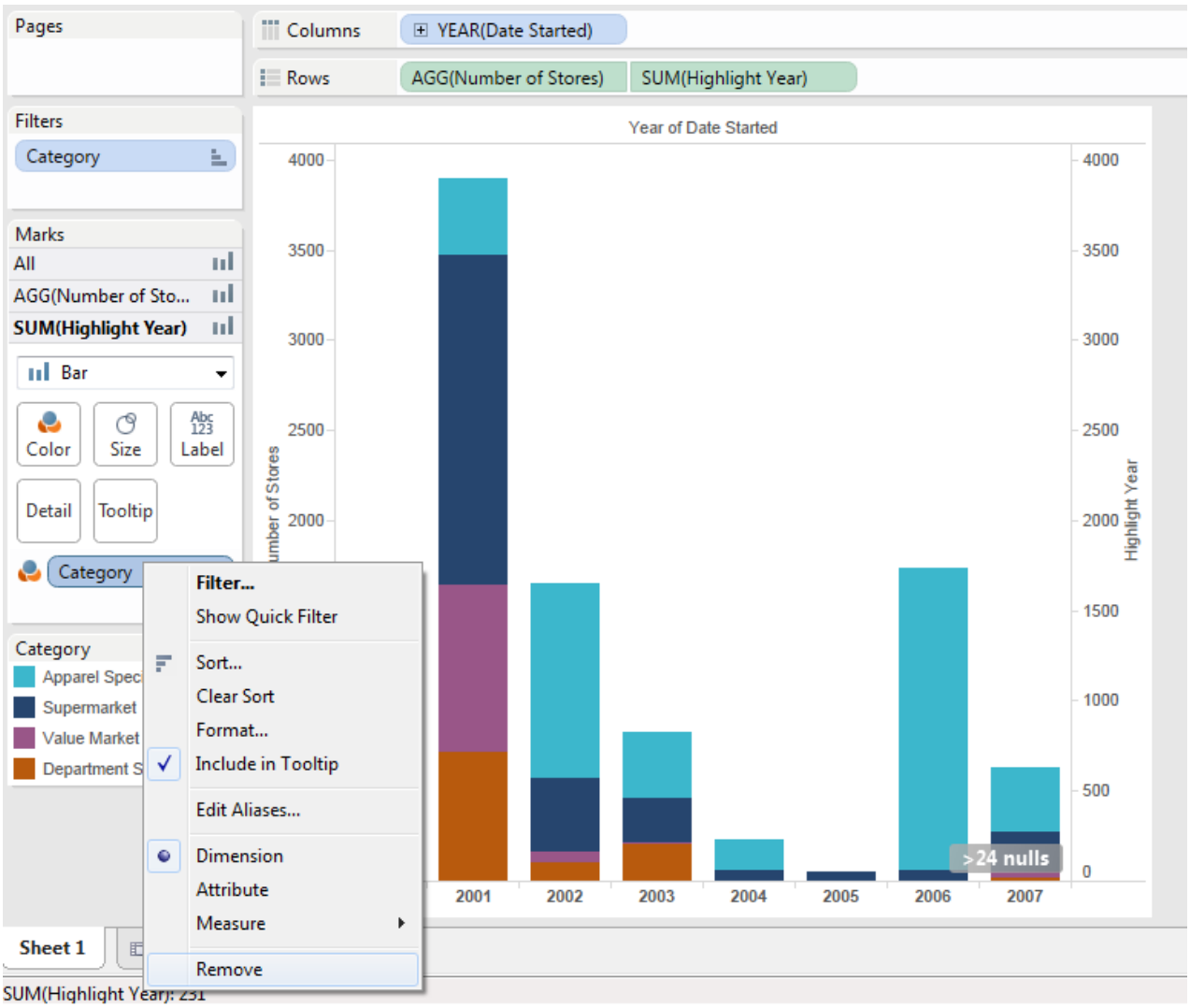
2.4. Make the Highlight Year row as a dual Axis



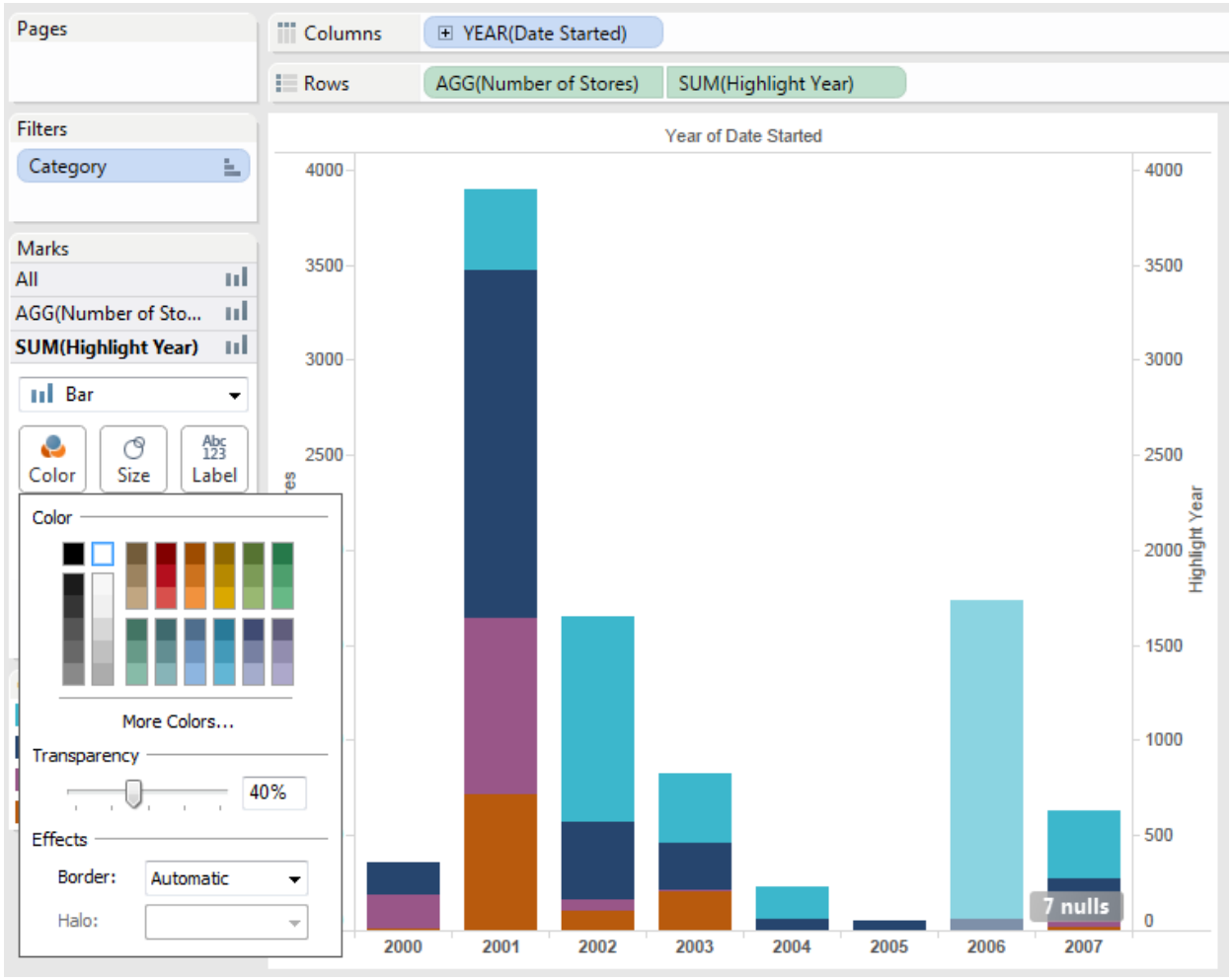
2.5. Right click on the right hand side axis and synchronize the Axis



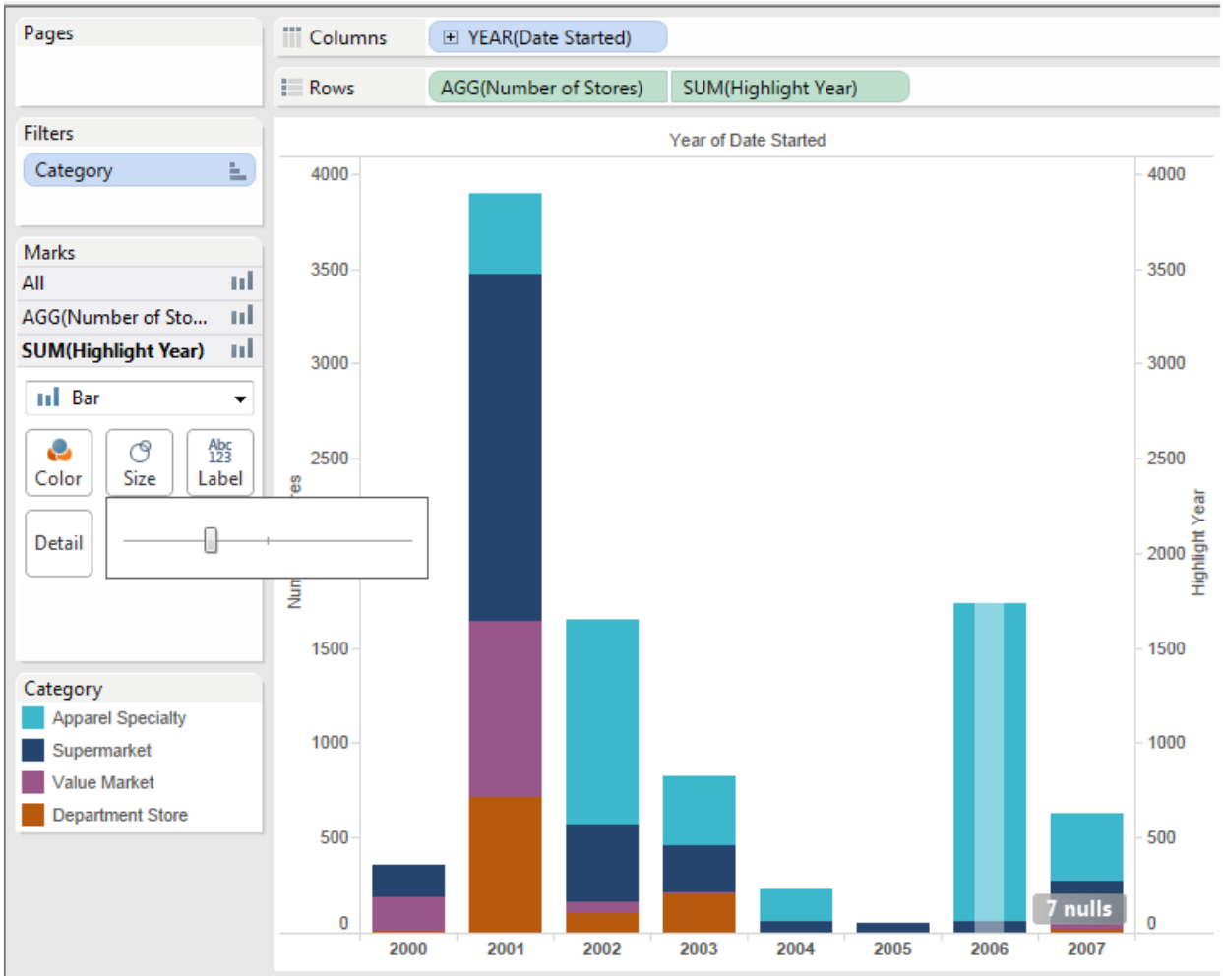
2.6.In the Sum (Highlight Year) marks; remove category from color



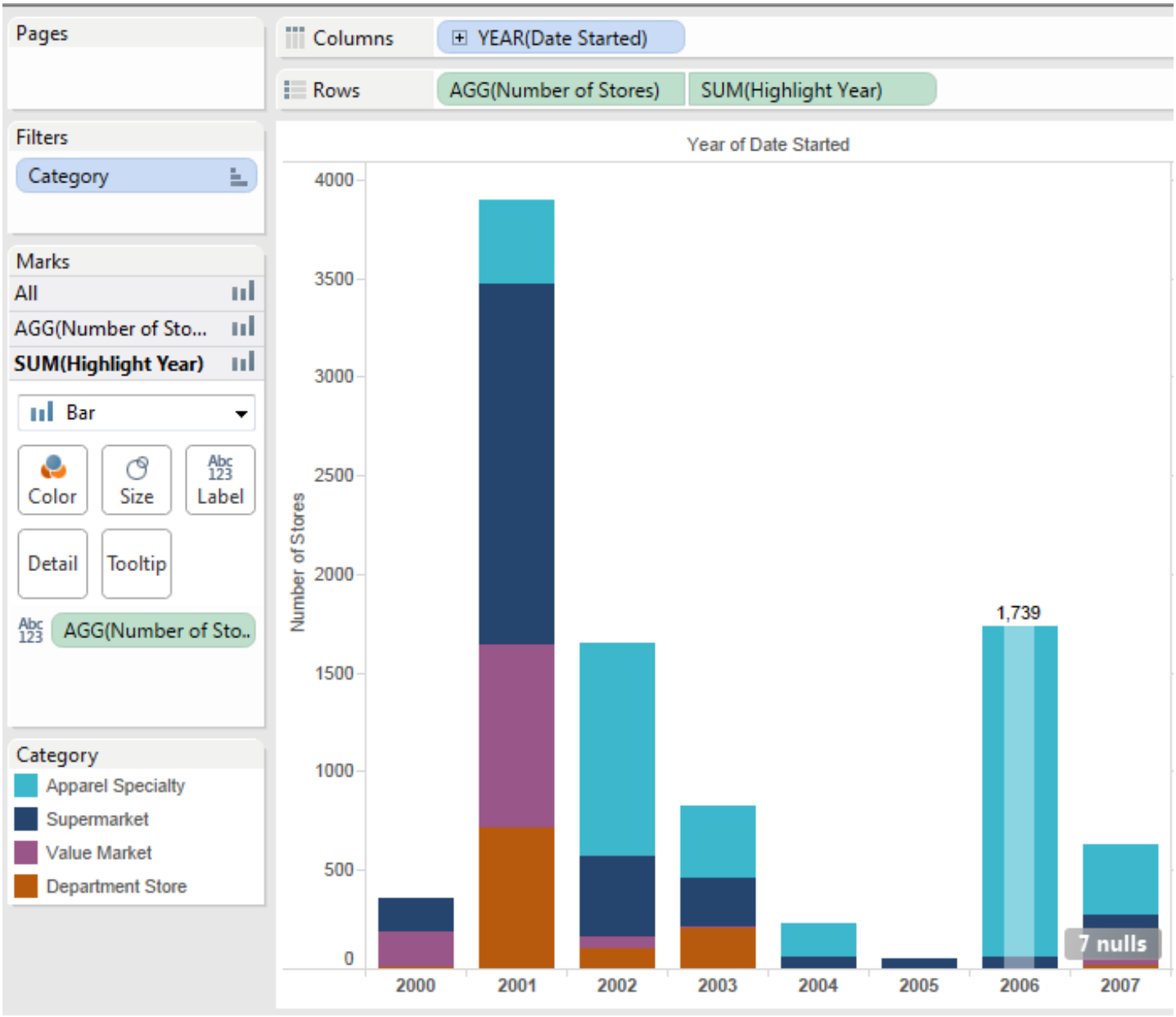
2.7. Change the color to white and transparency to 40%



2.8. Change the size to lesser size using the slider



2.9. Add the number of stores to the label



Resources:

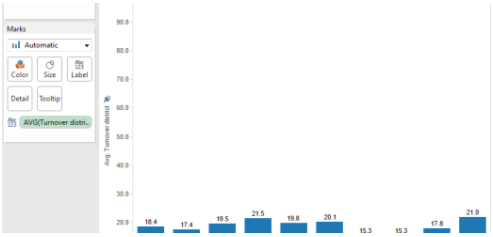
Initial tableau workbook Combination Chart (Right-click and “Save link as…”)

Original creation:

http://public.tableausoftware.com/shared/2QCDD7BPR?:display_count=no

Tableau exploration with Portfolio Management System

By Arunkumar Navaneethan



Gathered and analyzed requirements for a portfolio management system following an iterative development approach. Created an OLTP in MySQL and star schema, OLAP for the data warehouse, followed the standard design convention. The OLAP was created accommodating slowly changing dimensions recording historical data of employees and time stamped status of project information. Create facts and dimension table in MYSQL using complex SQL join statements on the OLTP tables. The dimensions and facts are created with indexes on foreign keys to enhance query performances. Created a view of the MYSQL data and the data extracts were loaded to the tableau server for query performance.

The tableau dashboard and workbooks are hosted in the tableau public server. Created calculated fields to find the margin based on the revenue gained and expense incurred. Structured the data as hierarchies with portfolio as the highest level followed by application and project details. Created groups for the application data to do category based analysis of the data. Created sets to display project data based on conditions like the top 3 projects for the given portfolio.

Created parameters to set working hours goals by checking if the working hours is more than a certain definite average. Created animated visualization to understand the growth of margin with respect to revenue and expense over time using the pages card. Created symbol maps to understand the geo-spatial distribution of projects with custom symbols made to represent the application category. Manually organized the dimensions and measures like moving the surrogate keys, date, descriptions, flags and categories as dimensions.

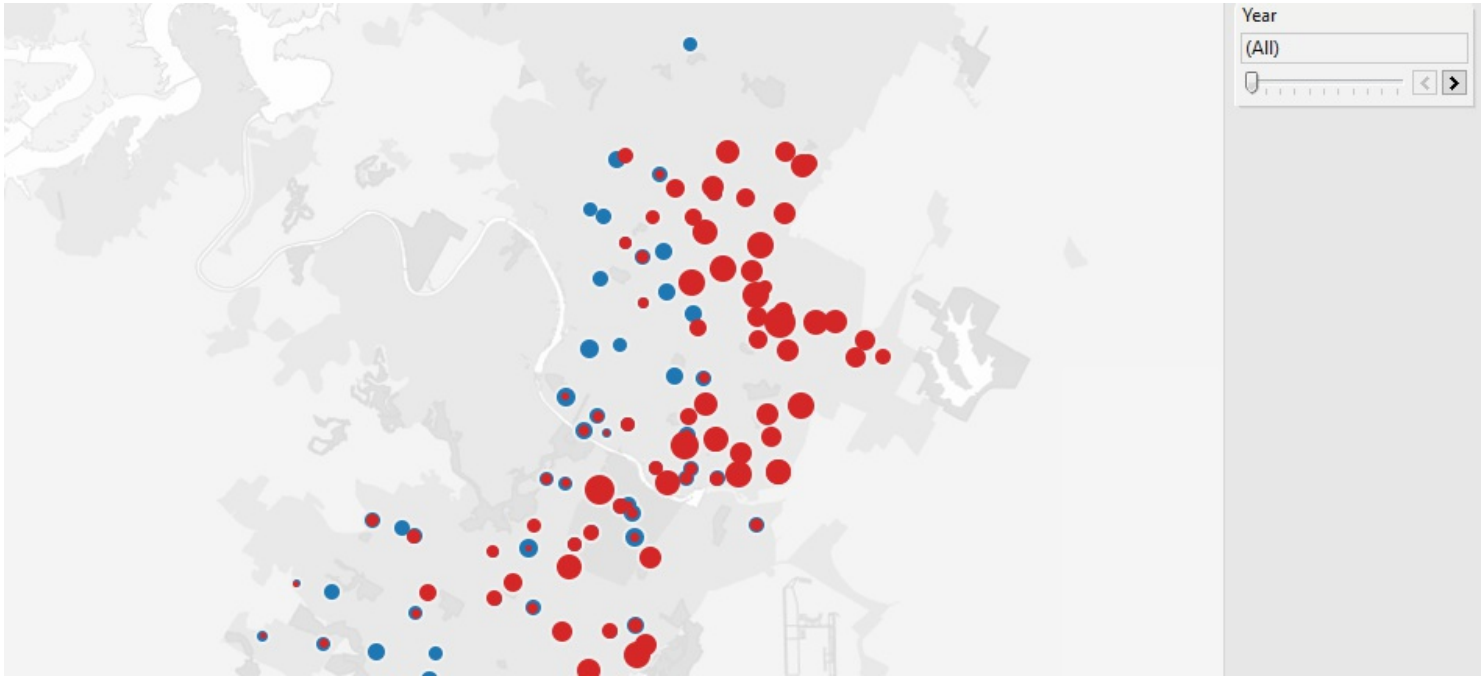
Customized the tooltips and text to end user understandable format, the numerical measure was changed to its appropriate formats like representing the revenue, expense and margins as currency values. Created dashboards in Tableau to analyze the revenue, expense over various attributes of employee, project, time and location dimension. The dashboard is built with floating objects and capitalizing the various dashboard features like URL, images, web integration's. The portfolio data is represented as tree map based on the margin, the detailed application data is represented as column graph with highlight table distributing the margin across employee roles. The detailed cross tab data of each employee is provided to understand the underlying data across attributes. Created the appropriate filter for project level and employee level data. The project level filter was single select and employee level filter was a wild card search. Experimenting with social media analytics to provide application related data from twitter REST API and streaming API.

The next step is to create a portal style dashboard using the existing dashboard and workbooks. Exploring tableau's javascript API features to implement this.

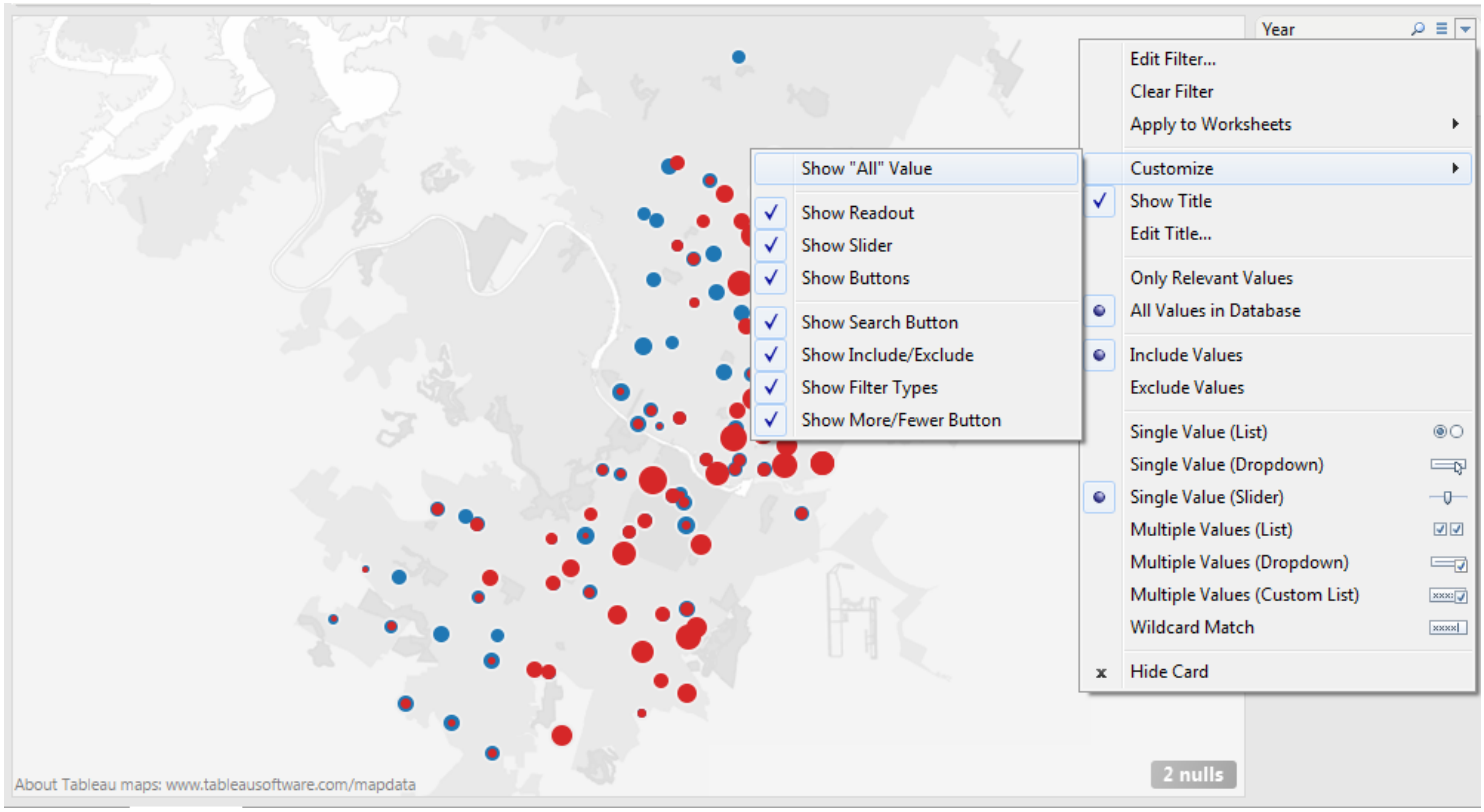
Tableau filter excluding All

By Arunkumar Navaneethan

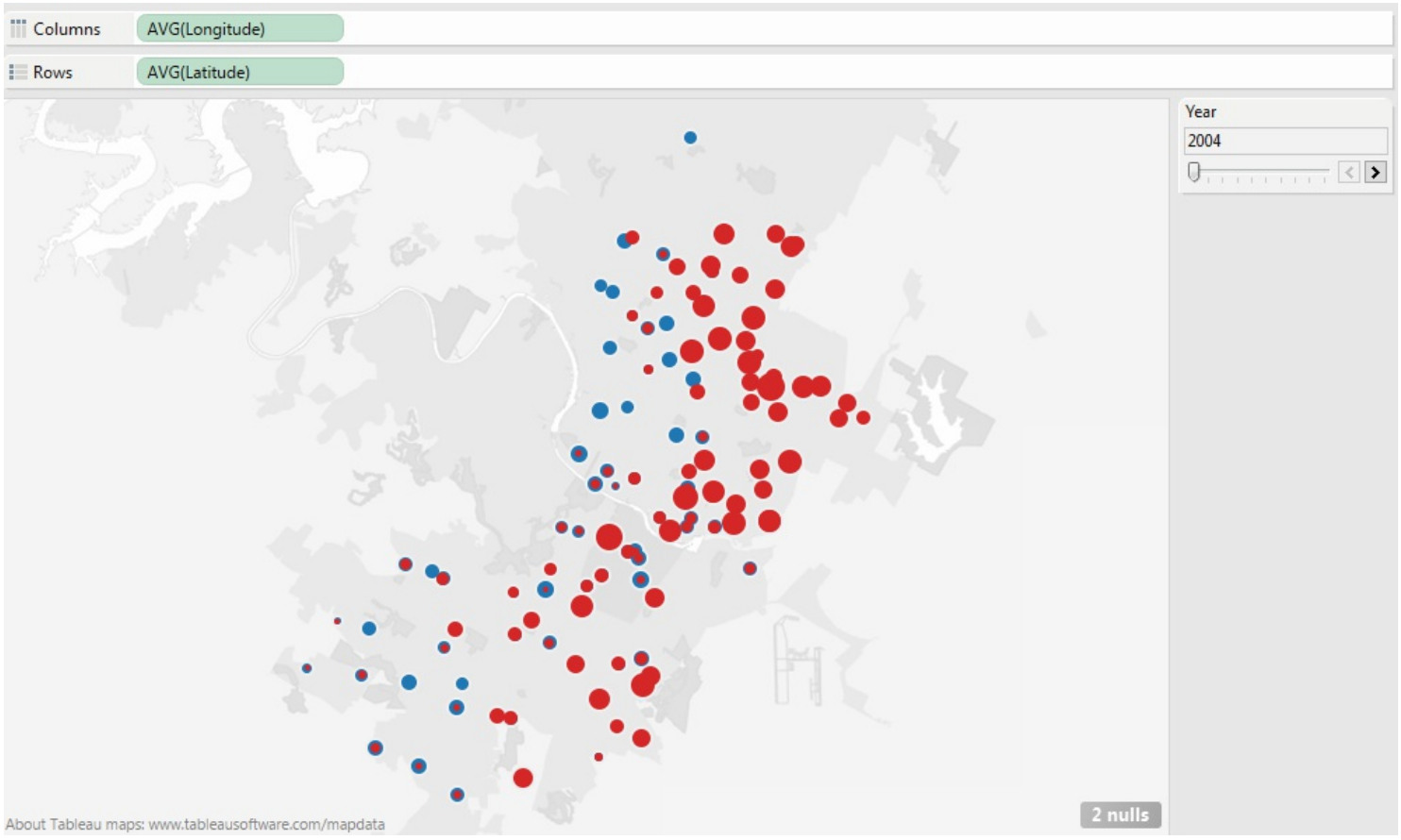
The default filter will start from “All”



Change the custom filter settings



Filter excluding “All”



My Use-Case:

When a data-set consist of data across multiple years or multiple categorical separation. The visualization is expected to display diversity across years or categories and not the aggregate of “All” the data. Then we can follow this procedure.

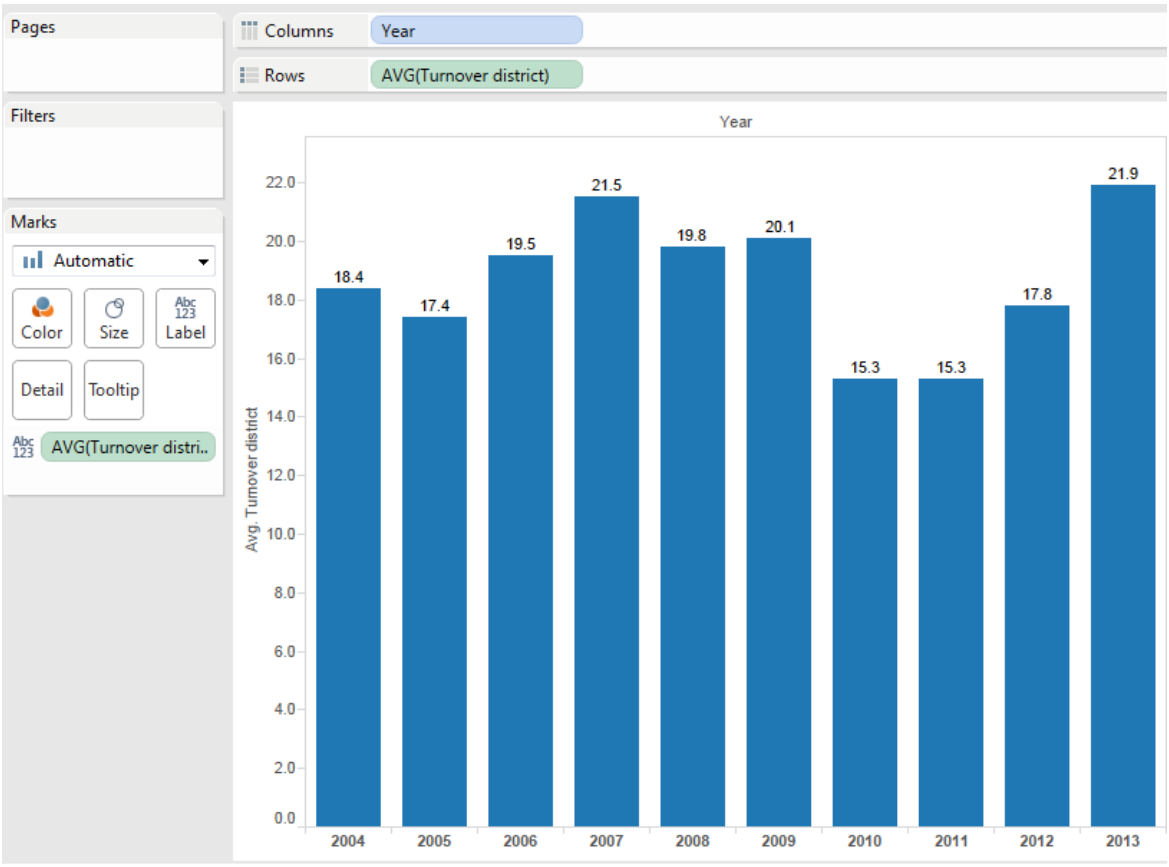
Tableau fixed axis length

By Arunkumar Navaneethan

My use case:

The underlying data is percentage turnover for different regions. The maximum value was around 21.

Tableau default axis length settings is automatic. It created a chart with a range of 0 – 23, but the percentage representation of the data is lost.

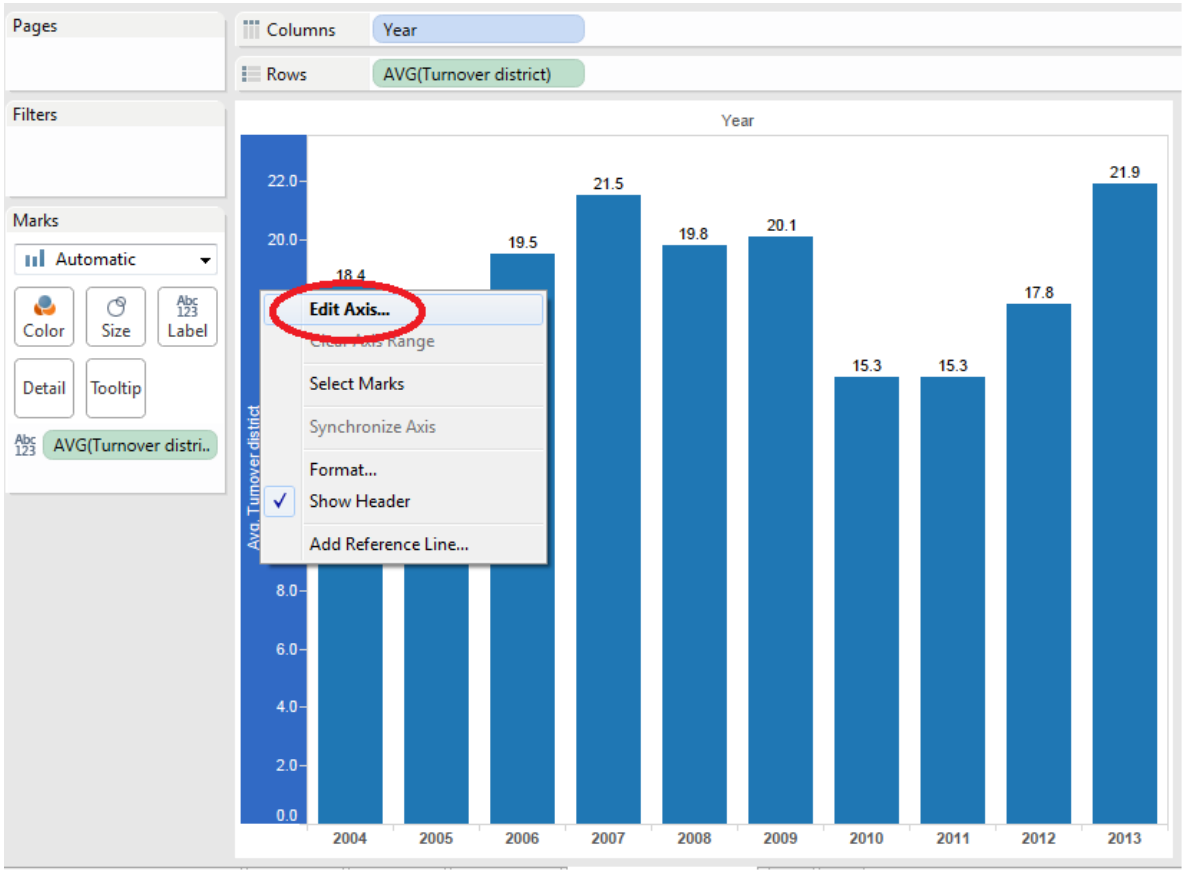


Before we start

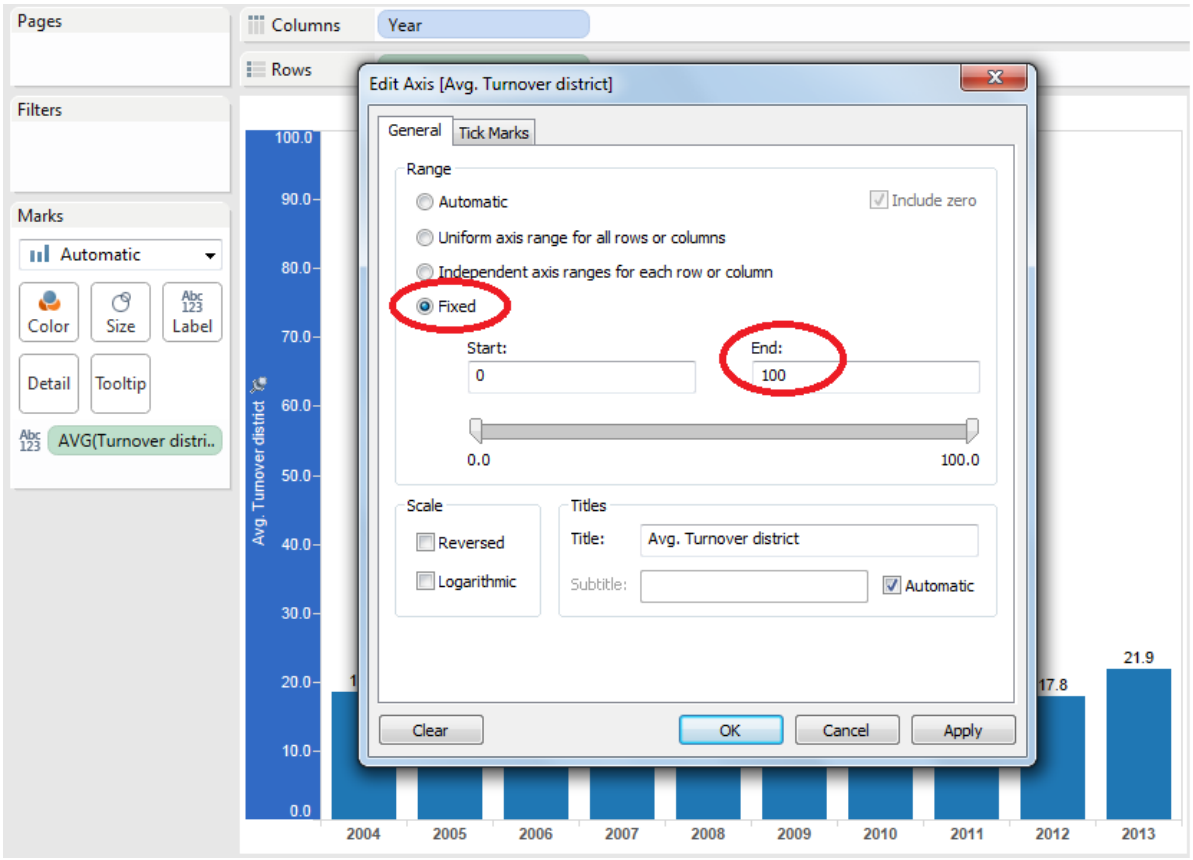
The axis setting can only be modified when the column or row value is continuous and not discrete.

Steps

1. Right click on the axis and select “Edit Axis”



2. Change the “Fixed” range to 0 – 100



Result

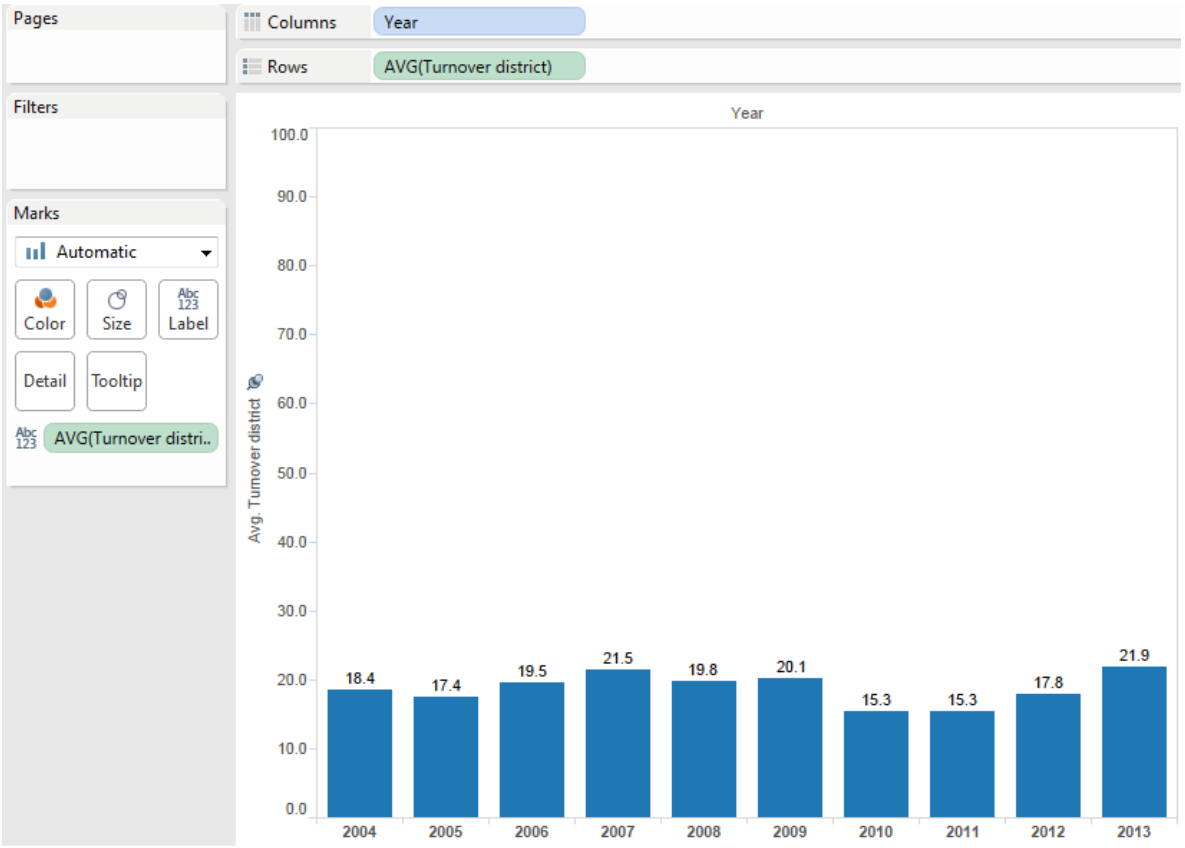


Tableau maps background colors

By Arunkumar Navaneethan

My use case:

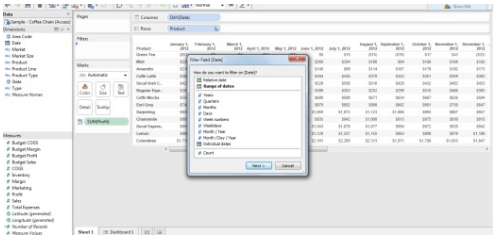
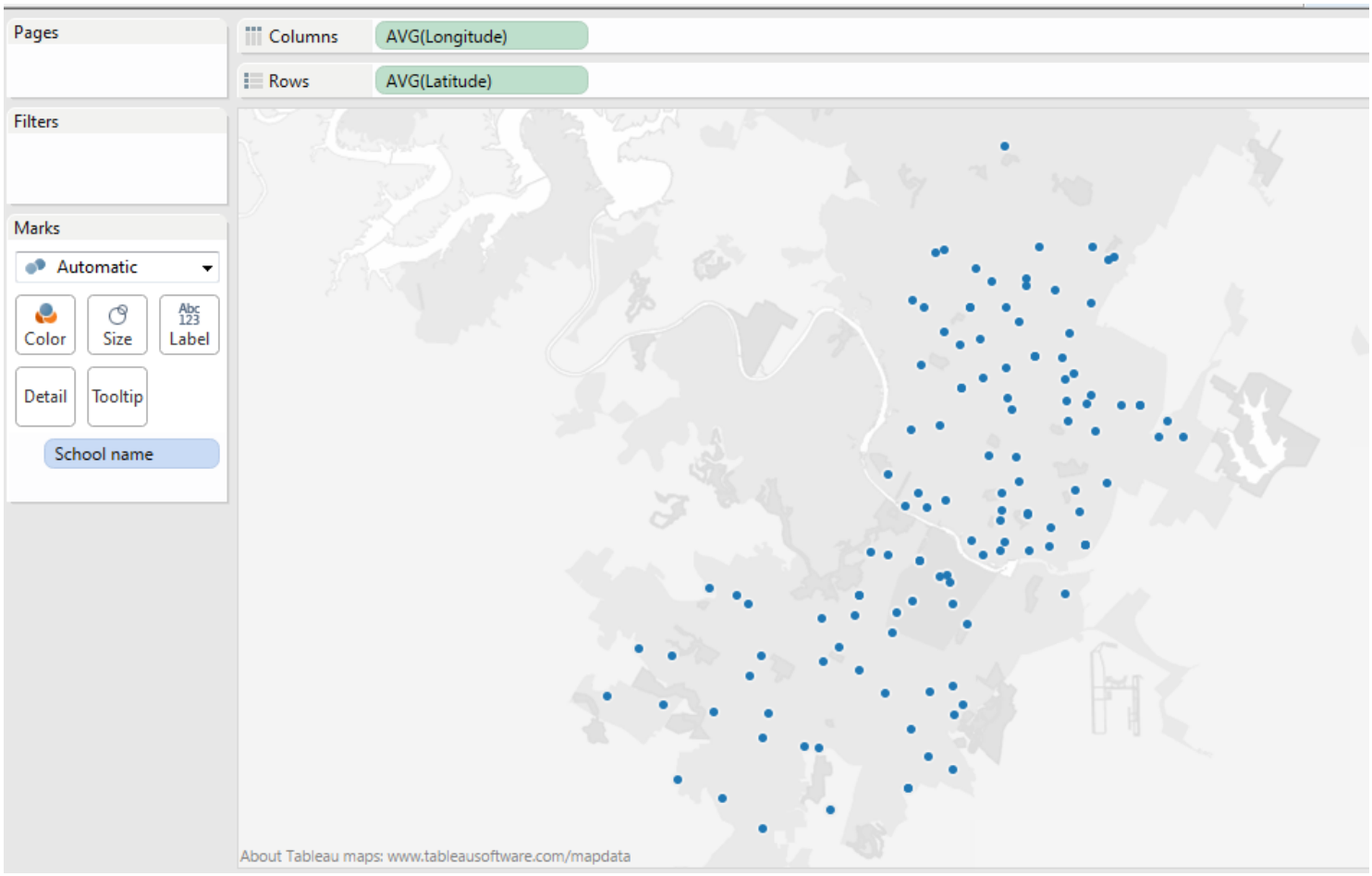
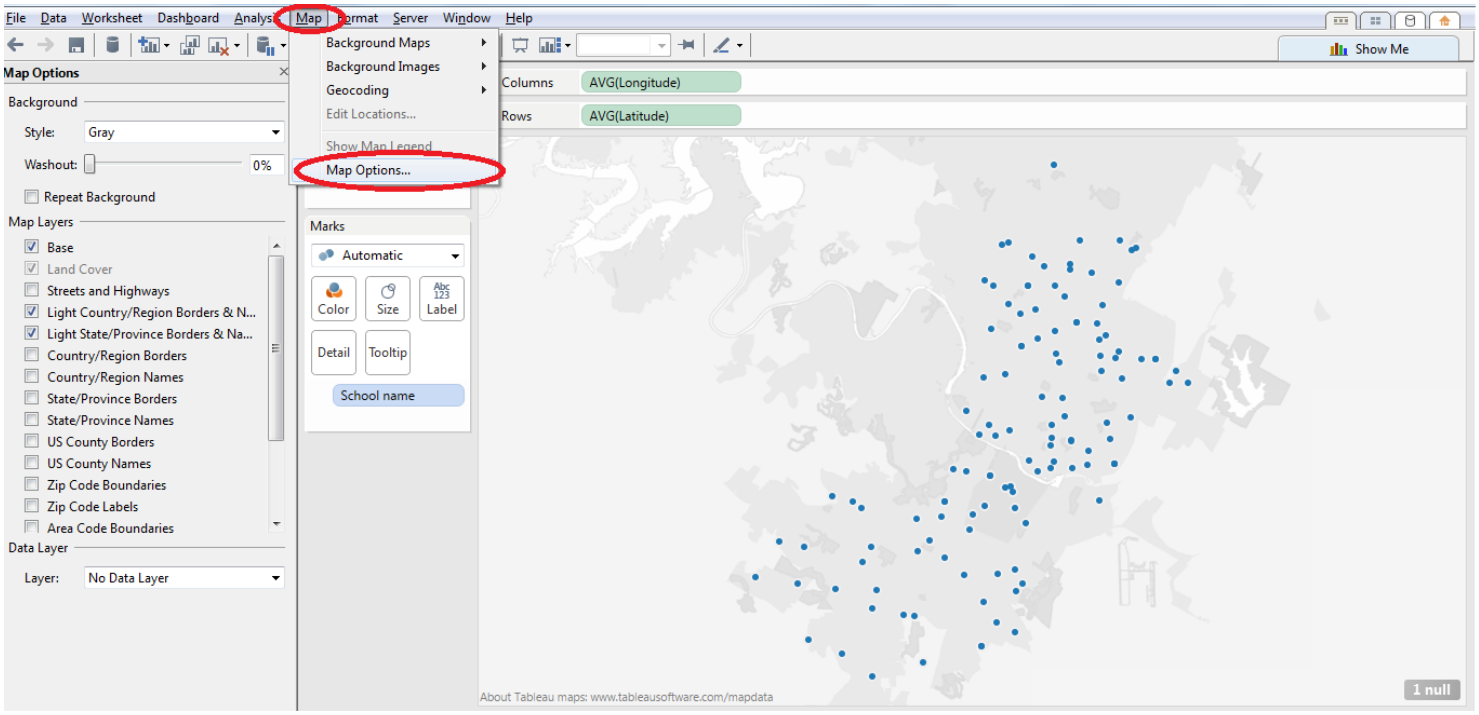


Tableau maps visualization for data from a Single state

I have data for teacher turnover in schools within Austin, TX (2004 – 2013). I used the user defined latitude and Longitude values, plotting school name as details.

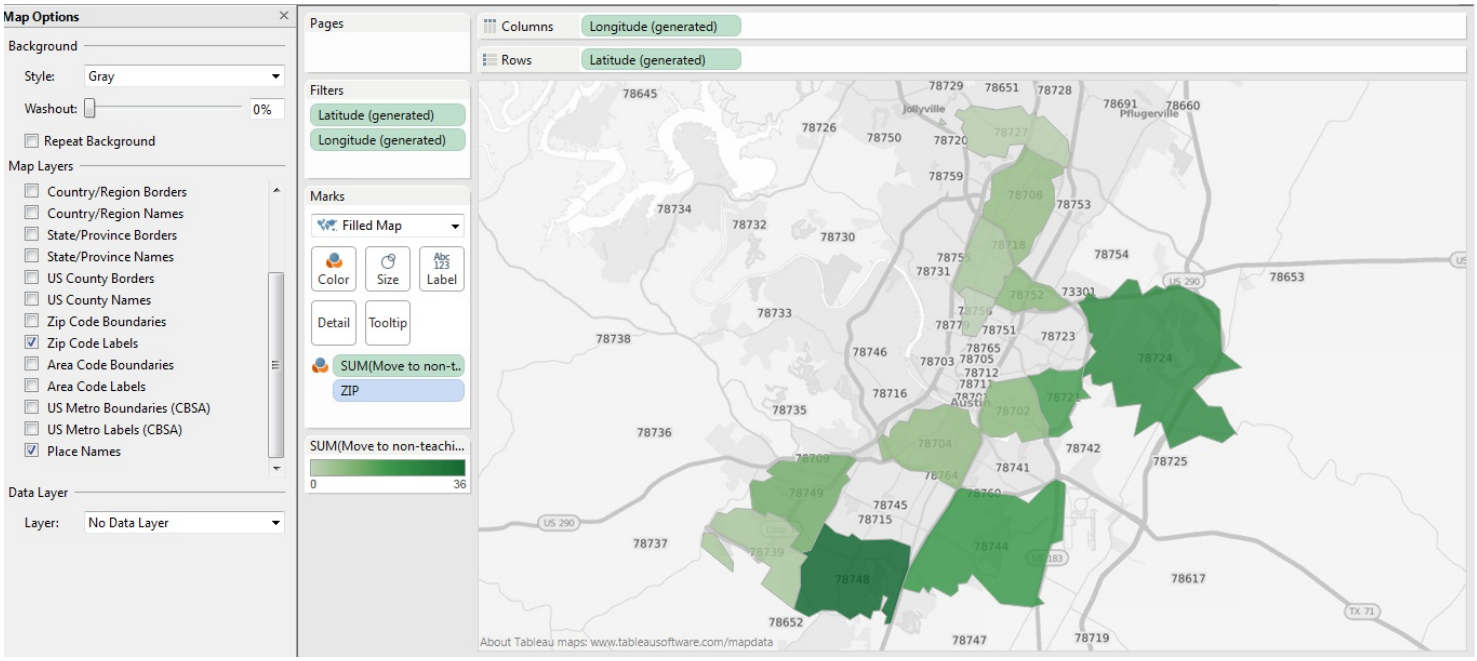


The map can be customized using the “Map Options” which is under “Map” tool bar.

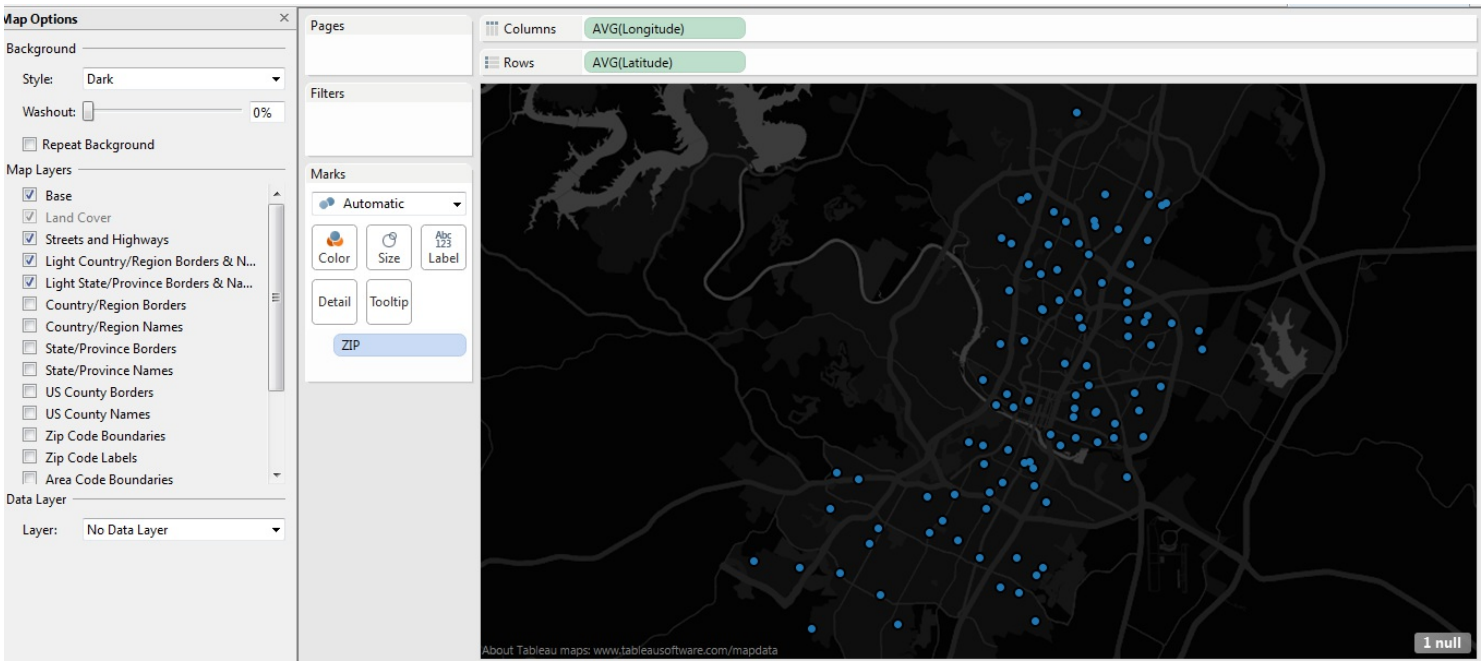


Background color options: gray, dark and normal

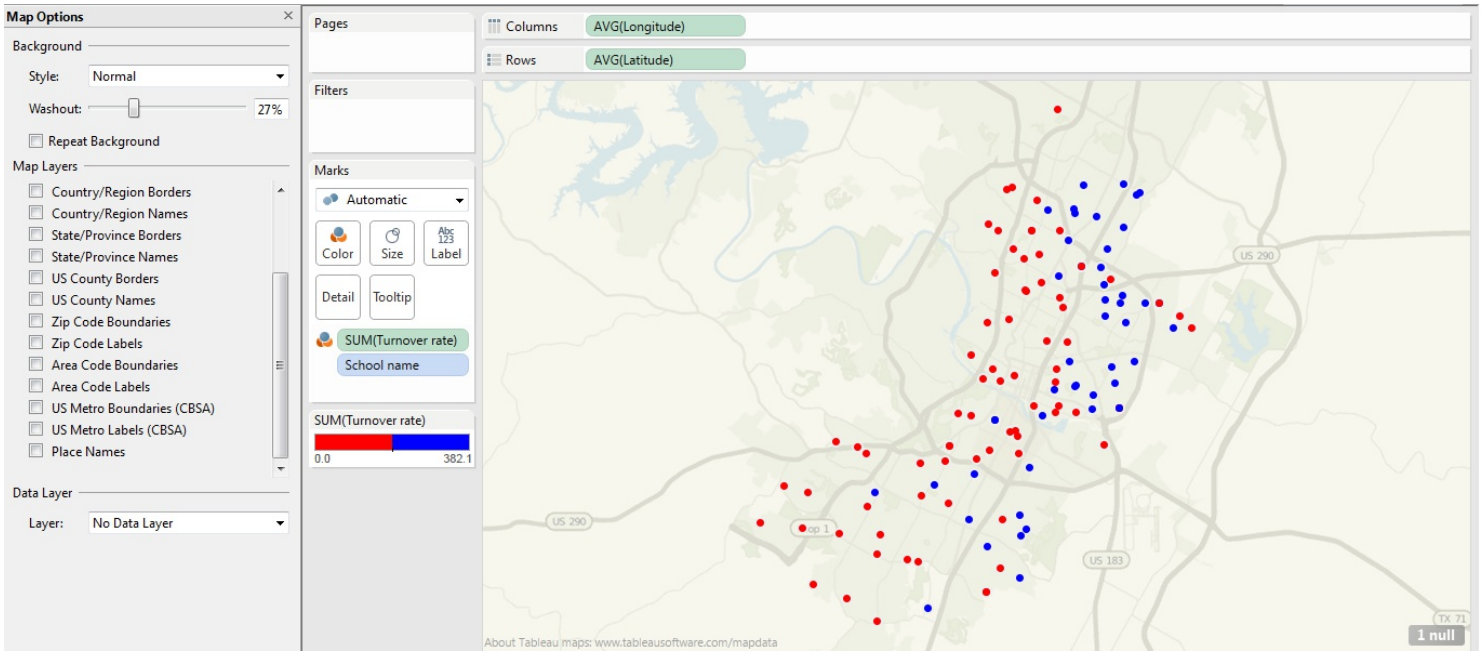
The default map background color is “Gray” it is typically suitable for symbol maps and filled maps. It is best to use when map layers like “zip code labels”, “place names” are displayed



The dark background is more suitable for the symbol maps.



The normal background is rarely used but it does highlight the geographical features like rivers and landscape natural.

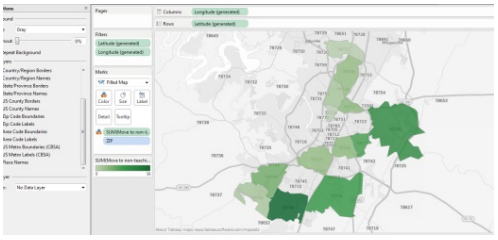


Data source:

http://www.mystatesman.com/interactive/aisd-teacher-turnover/

Tableau Mobility – Approach

By Arunkumar Navaneethan



In Tableau, you do not have to do anything special to make a dashboard mobile. Simply publish to Tableau Server like you always have, and Tableau will detect if you're using the Mobile app. You even get the native touch experience if you go to Tableau Server from your mobile browser, without the app at all.

Ease of use is the single most important aspect of mobile business intelligence. When you're on the go, you need to be able to get to the data you need with a few taps.

Filters: Tableau's controls such as filters, parameters, sliders, scrolling, and zoom & pan, are specially built to interact with your fingers. For example, tapping a filter pops a large, touch-optimized quick filter. And for long filters, there is scrolling inside the filter.

Views: Views themselves are touch-optimized with dynamic scrolling. Simply swipe to scroll through a long customers list, for example. Or pinch & zoom in a map.

The two major approaches for Mobility in Tableau

Security

Mobile business intelligence must be secure. With the Tableau mobile solution, security and metadata continues to be managed by Tableau Server. This means you can enforce your existing security protocols and integrate with ActiveDirectory via Tableau Server.

And if an employee loses their iPad or Android tablet, simply disable their Tableau Server account and give them a new one. No data other than descriptive data about a workbook (like the publisher, data modified and name) are stored on the device, so you can keep your data secure even while it's mobile.

Limitation

Write to a smaller form factor ["author-once, distribute everywhere"]

Smaller screen? You've got to use your space better. Place the most important view at the top of your dashboard. Be stingy with legends and filters, and for that matter views– limit your dashboard to 2-4 important views. 10 views in a dashboard doesn't work on the go!

Reference

1. "5 Best Practices for Mobile Business Intelligence." *Tableau Software* 1 (2011). http://www.tableausoftware.com/sites/default/files/whitepapers/whitepaper_5-best-practices-mobile-bi.pdf (accessed April 8, 2014).

Tableau parameter as a filter

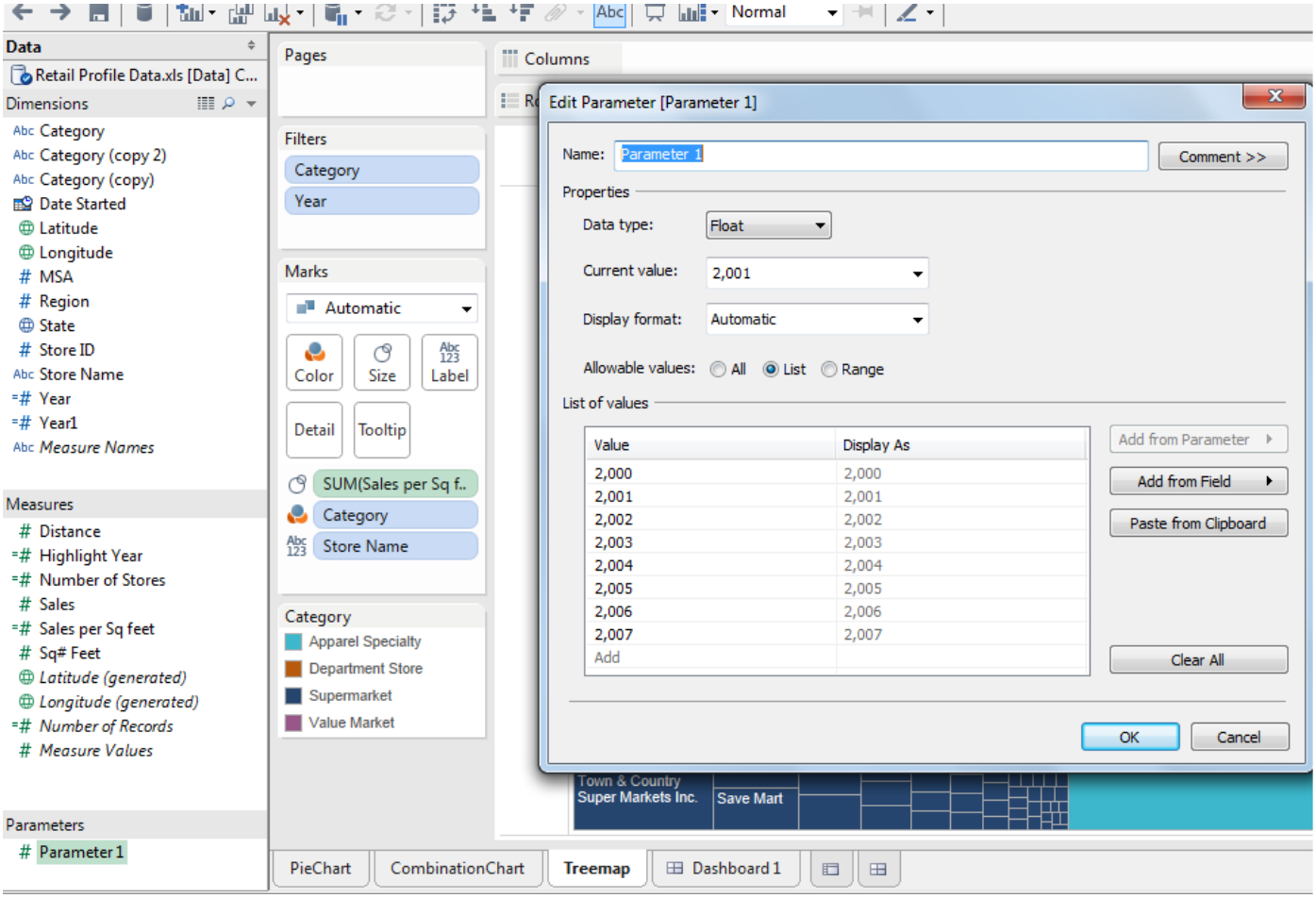
By Arunkumar Navaneethan

My Use-case

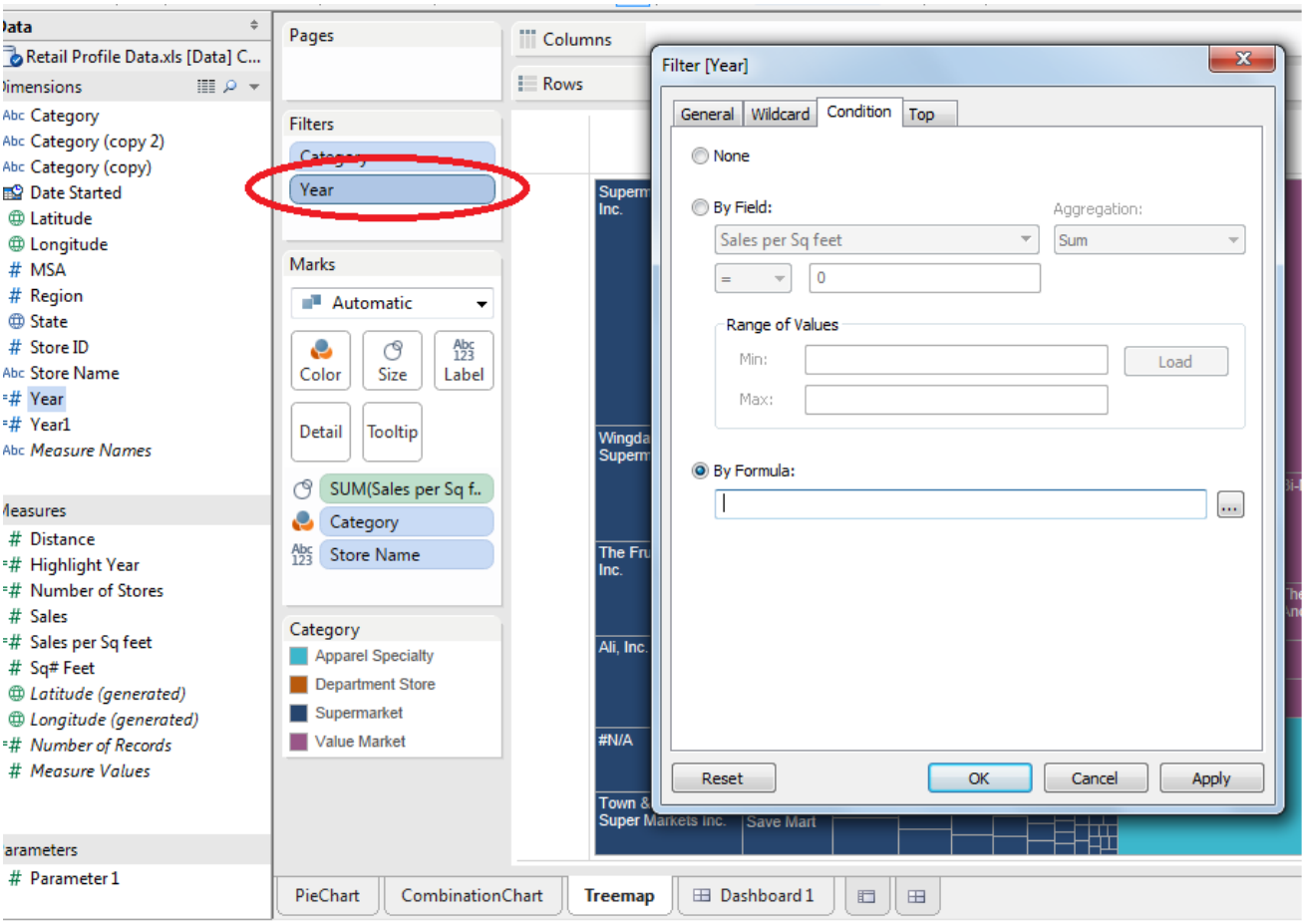
They are three different workbooks with data across years. Must link the workbooks together based on user selection of a year using parameters.

Steps

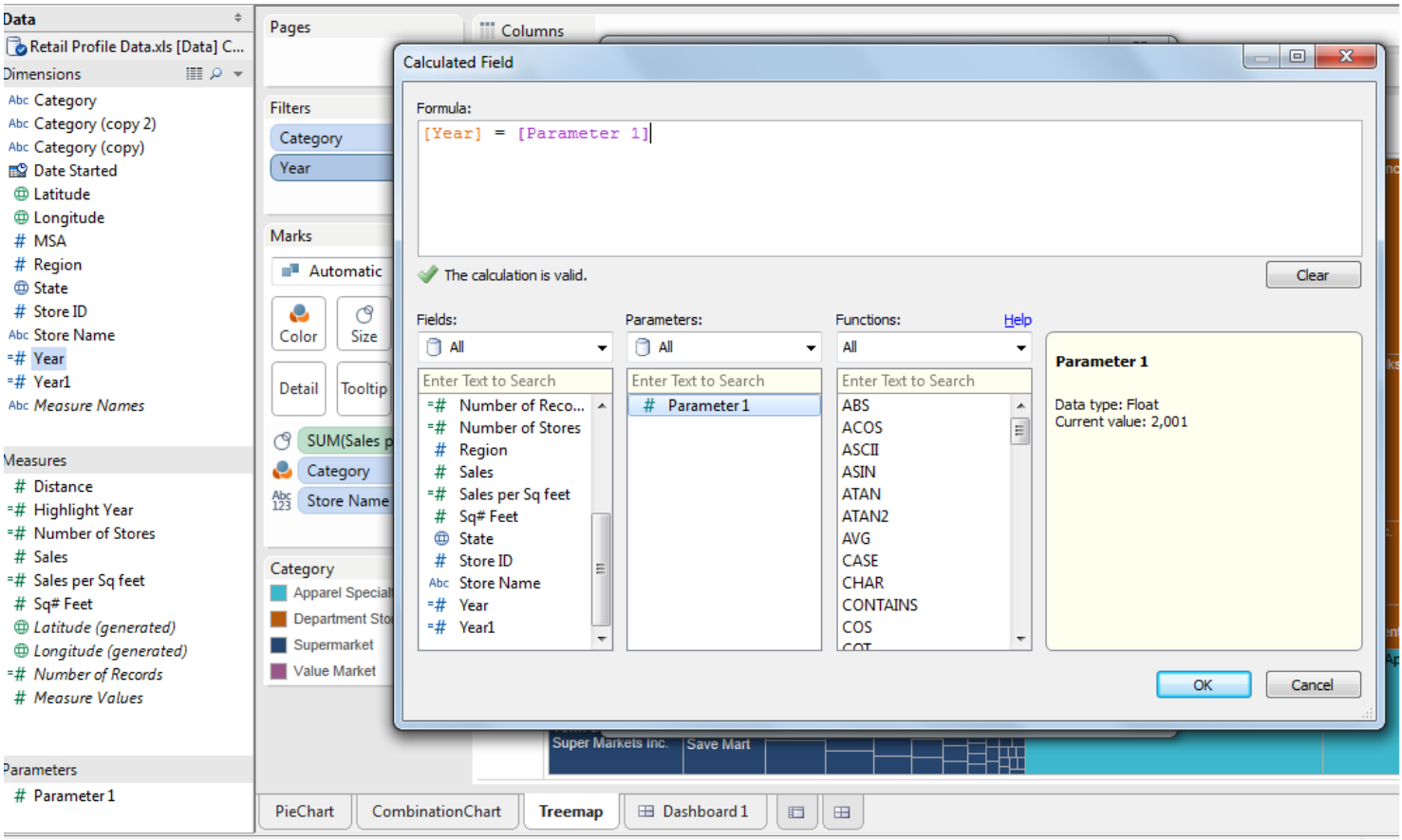
1. Create the Year Parameter



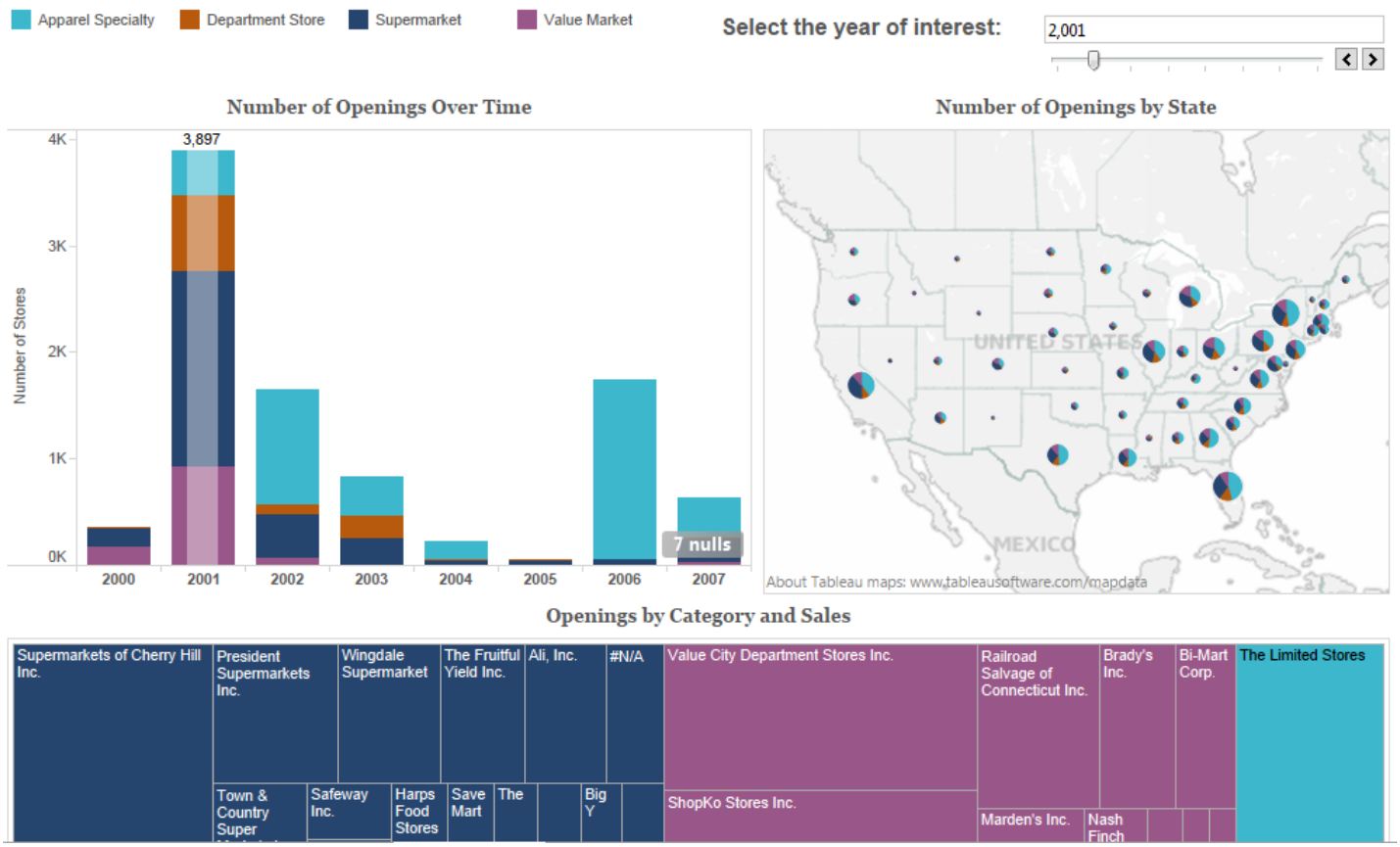
2. Drag the year dimension to filter box and open the condition tab



3. Use the formula ([Year] = [Parameter 1])



4. Follow the step 2,3 for the other workbooks



Resources

Initial tableau dashboard Parameter filter (Right-click and “Save link as…”)

Original Creation

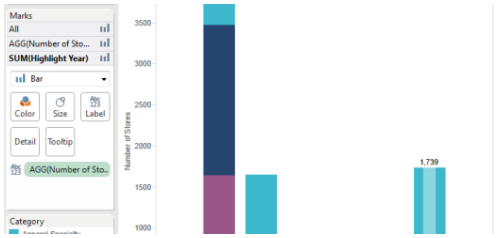
http://public.tableausoftware.com/shared/2QCDD7BPR?:display_count=no

Tableau Pie chart on a symbol map

By Arunkumar Navaneethan

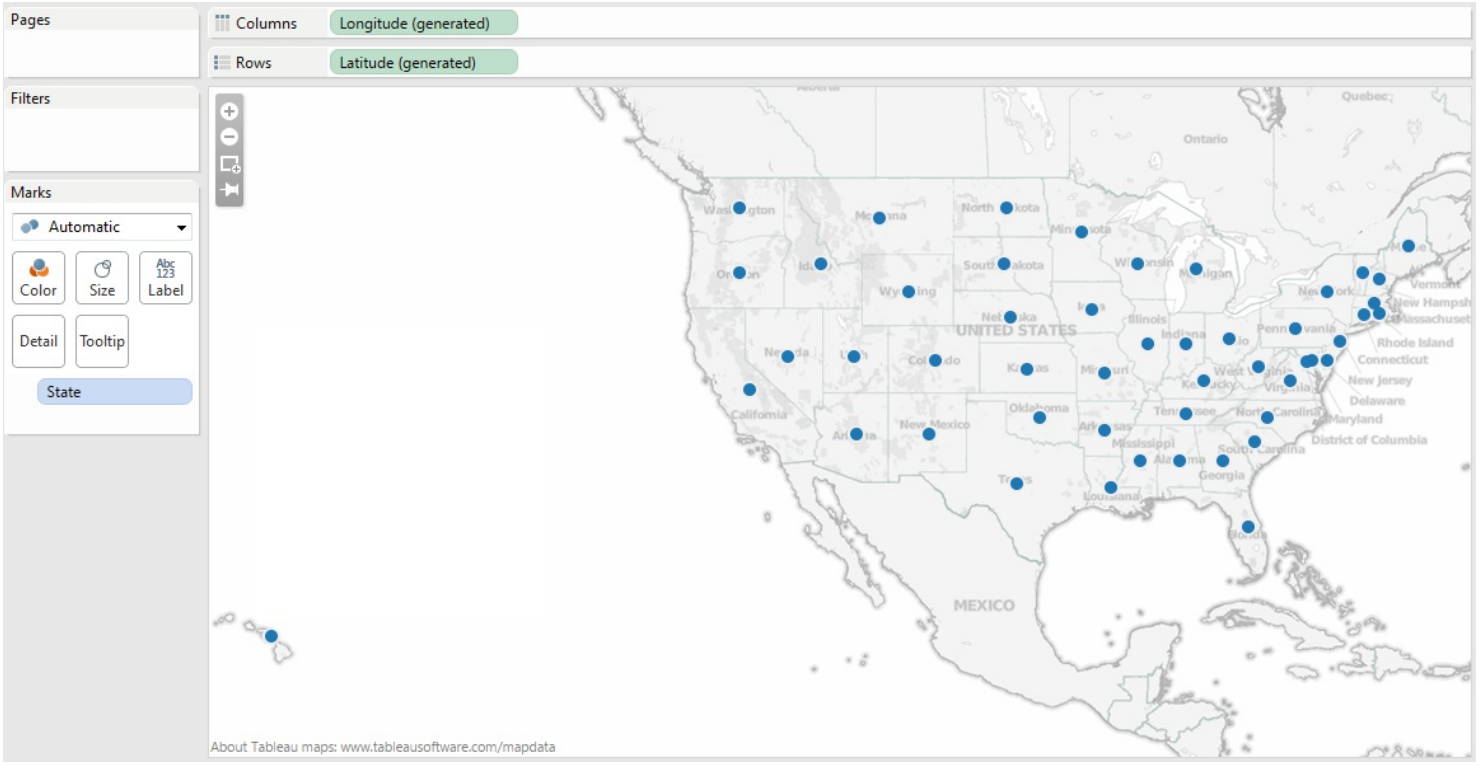
My Use-case

Display pie chart over every state, sized by the total number of stores. Size the slices of the pie chart by store categories. This will help us understand which states have more number of store opening and detailed by category.

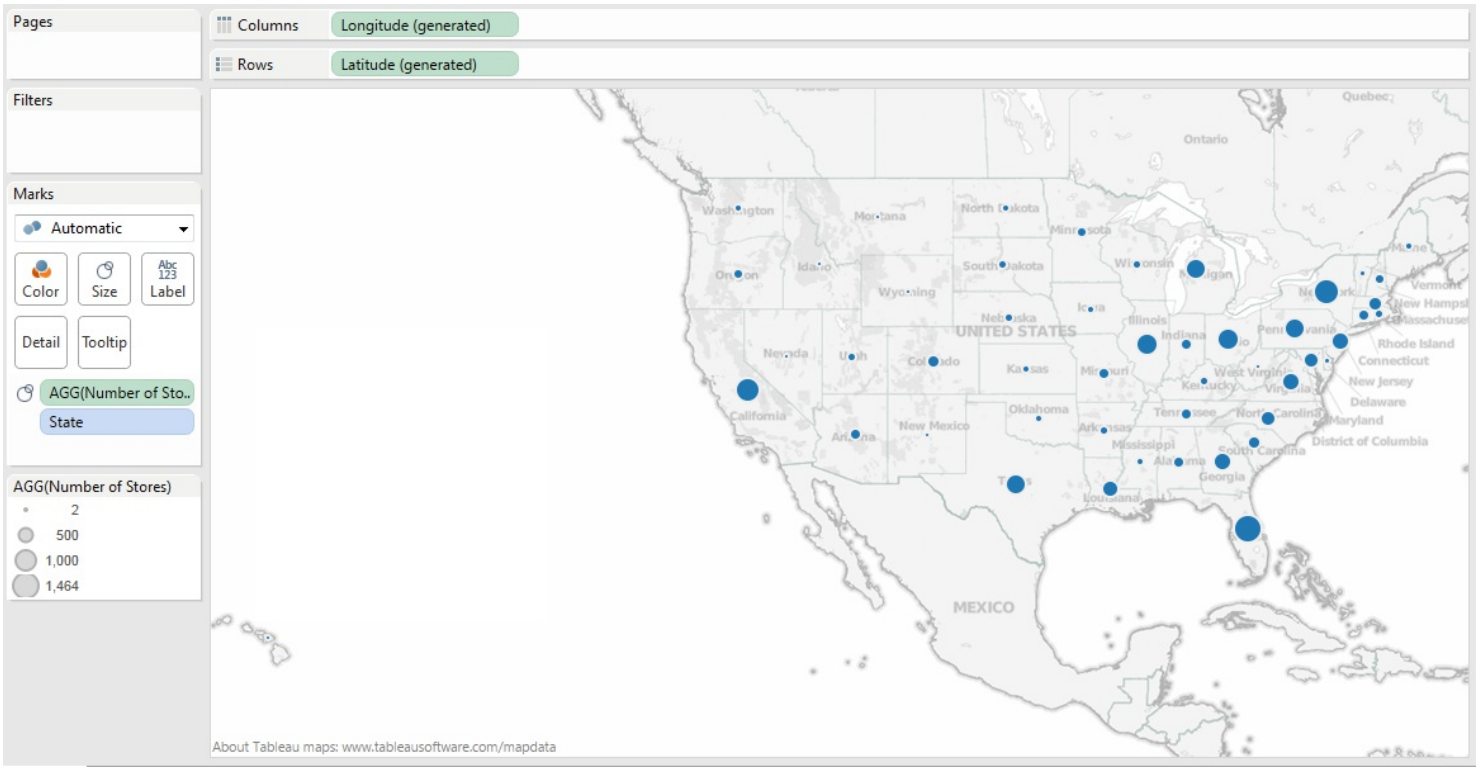


Steps

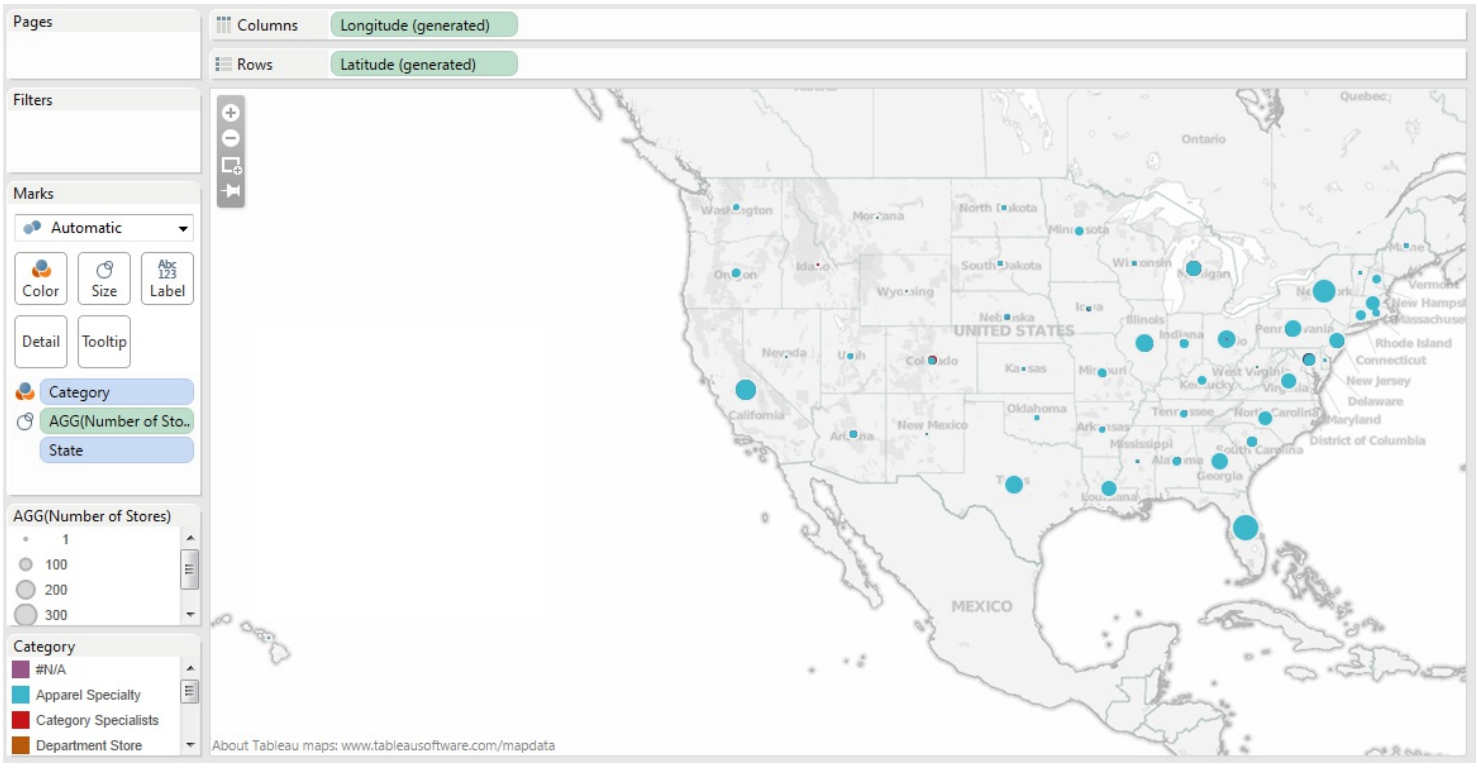
1. Add the state to the chart by double clicking on state dimension or add the state dimension to details and move the longitude measure to columns and latitude measure to rows.



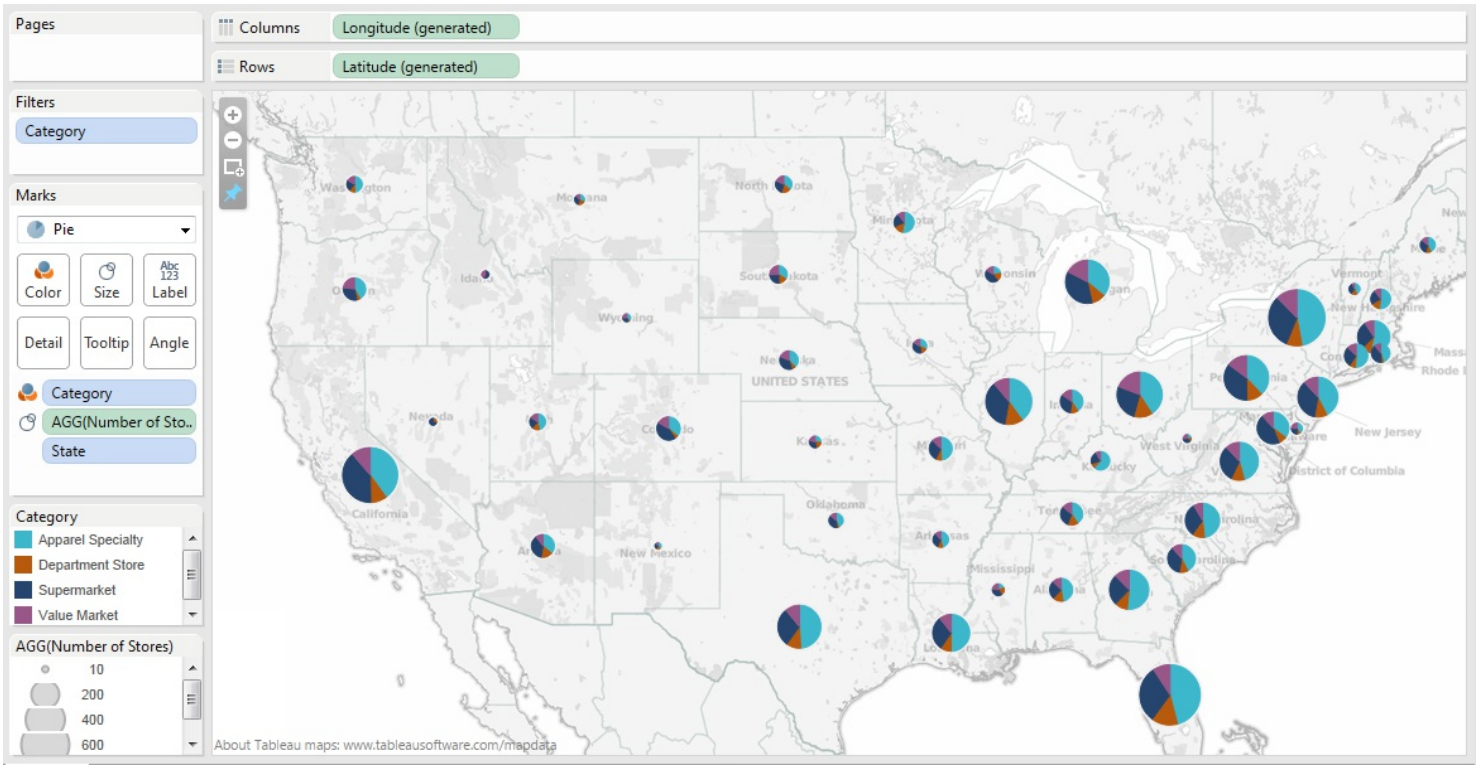
2. Move the Number of store measure to the size mark



Move the Category dimension to the color mark



3. Change the Mark type from automatic to Pie and use the size mark to resize the pie size. The map can be zoomed into a closer view if required



Resources

Initial tableau workbook Pie Chart (Right-click and “Save link as…”)

Original Creation

http://public.tableausoftware.com/shared/2QCDD7BPR?:display_count=no

Tableau URL action

By Arunkumar Navaneethan

My Use-case

We can carry out various external interaction using the URL action.

We can use this to redirect users to google search results, intranet sites, REST methods or predefined URLs

1. Google search

*https://www.google.com/#q=**<Store Name>** store*

2. Predefined URL's

<Store URL>

3. REST methods

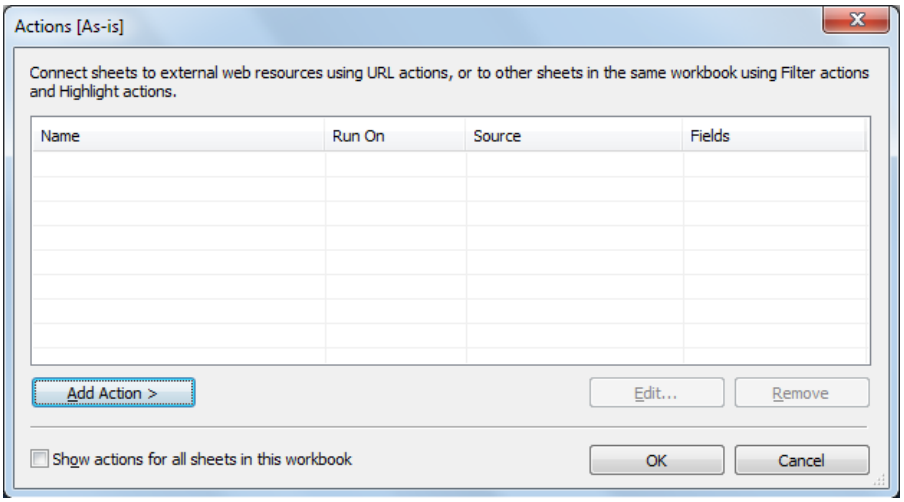
*https://www.arunethan.com/updateStoreInfo.jsp?StoreName=**<Store Name>***

4. Email with subject

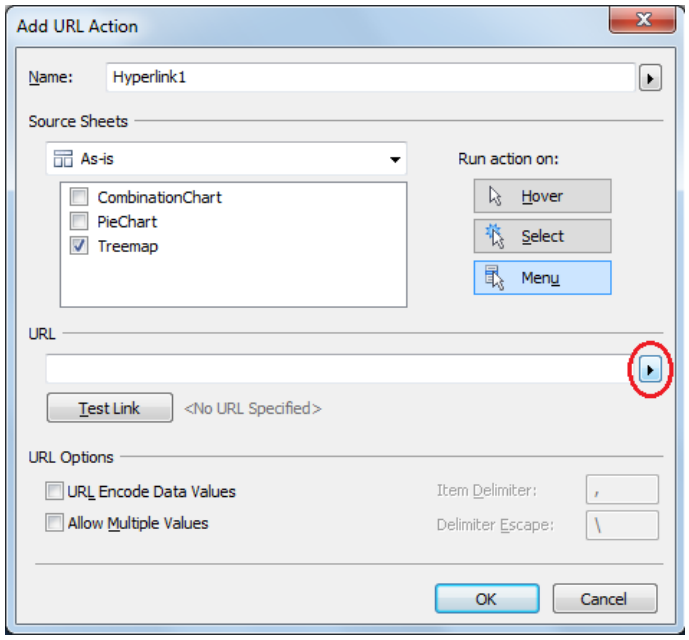
*mailto:info@arunethan.com?subject=Request for more information about **<Store Name>***

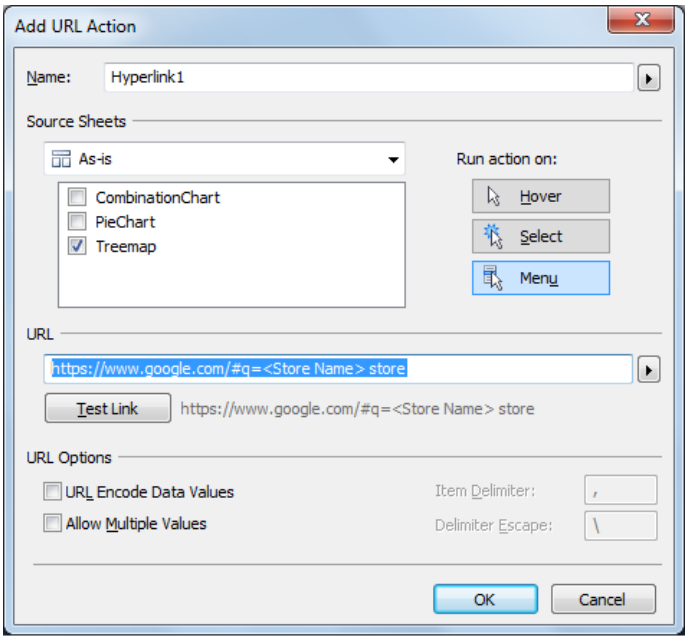
For a dashboard, open the Dashboard menu –> then Actions.. –> Add Action, button –> URL..

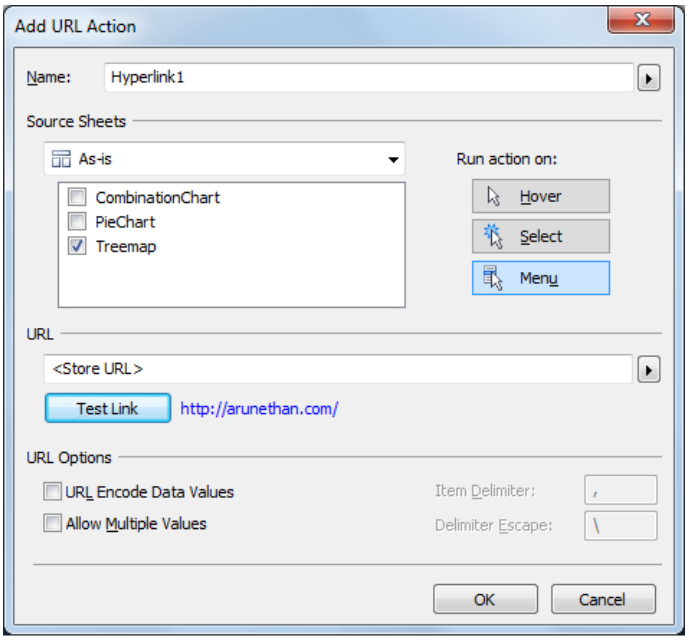
The URL can be created with values from tableau <dimension> or <measure> using the triangle button



Examples


Add URL Action
Name:
Hyperlink1
Source Sheets
As-is
CombinationChart
PieChart
Treemap
Run action on:
Hover
Select
Menu
URL
Test Link
<No URL Specified>
URL Options
URL Encode Data Values
Allow Multiple Values
Item Delimiter:
,
Delimiter Escape:
\
OK
Cancel


Add URL Action
Name:
Hyperlink1
Source Sheets
As-is
CombinationChart
PieChart
Treemap
Run action on:
Hover
Select
Menu
URL
https://www.google.com/#q=<Store Name> store
Test Link
https://www.google.com/#q=<Store Name> store
URL Options
URL Encode Data Values
Allow Multiple Values
Item Delimiter:
,
Delimiter Escape:
\
OK
Cancel


Add URL Action
Name:
Hyperlink1
Source Sheets
As-is
CombinationChart
PieChart
Treemap
Run action on:
Hover
Select
Menu
URL
<Store URL>
Test Link
http://arunethan.com/
URL Options
URL Encode Data Values
Allow Multiple Values
Item Delimiter:
,
Delimiter Escape:
\
OK
Cancel

Using Range/Relative filter for a discrete Date field in Tableau

By Arunkumar Navaneethan

Different Date filters

Relative Date - This filter helps to analysis the data with respect to year, quarter, month, weeks, days relative to the current date or an user created anchor date.

Range of Dates – If the data has to be analyzed based on certain user defined interval

Start Date and End Date – End date is helpful for As-of Date analysis

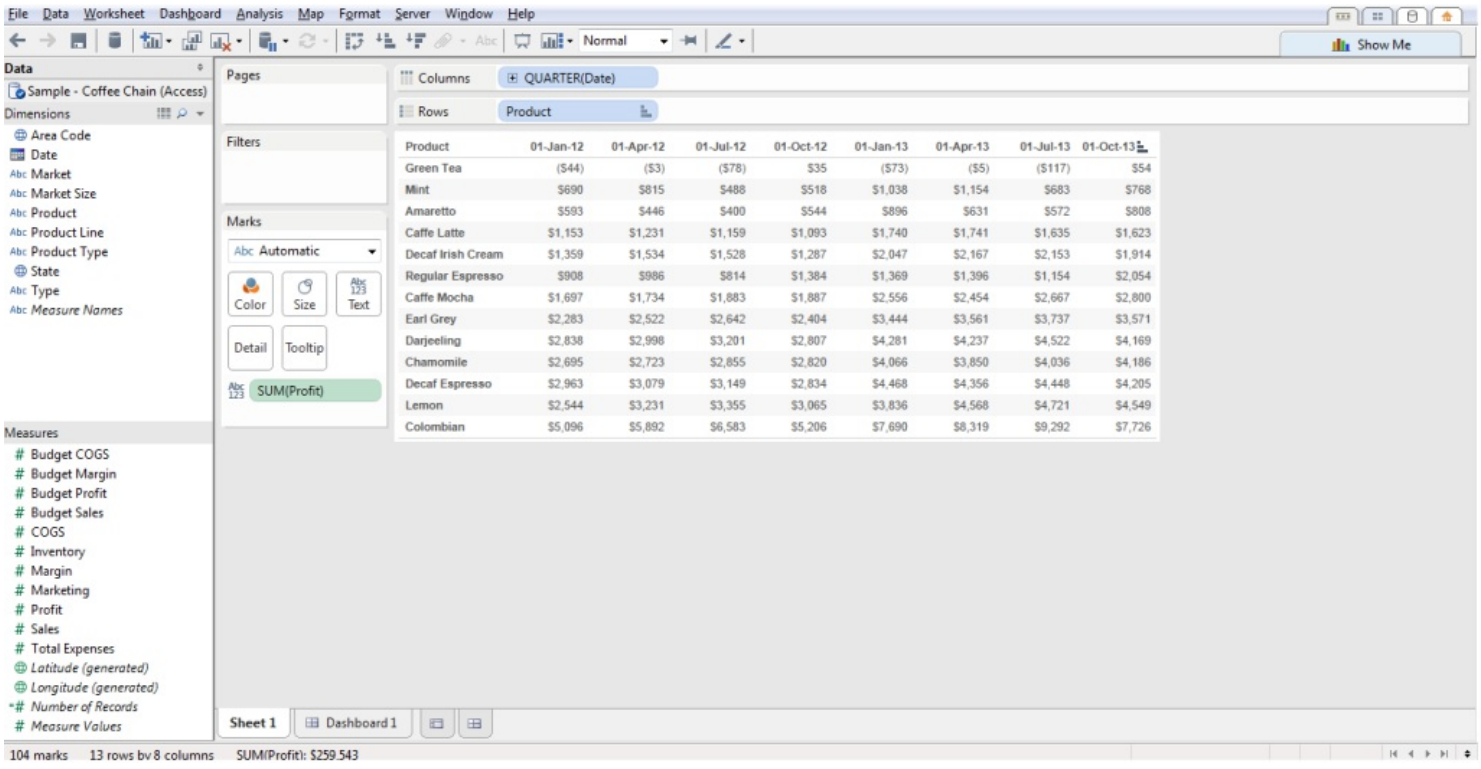
Browse Period – Data can be analyzed on various date granularity

My Use-case

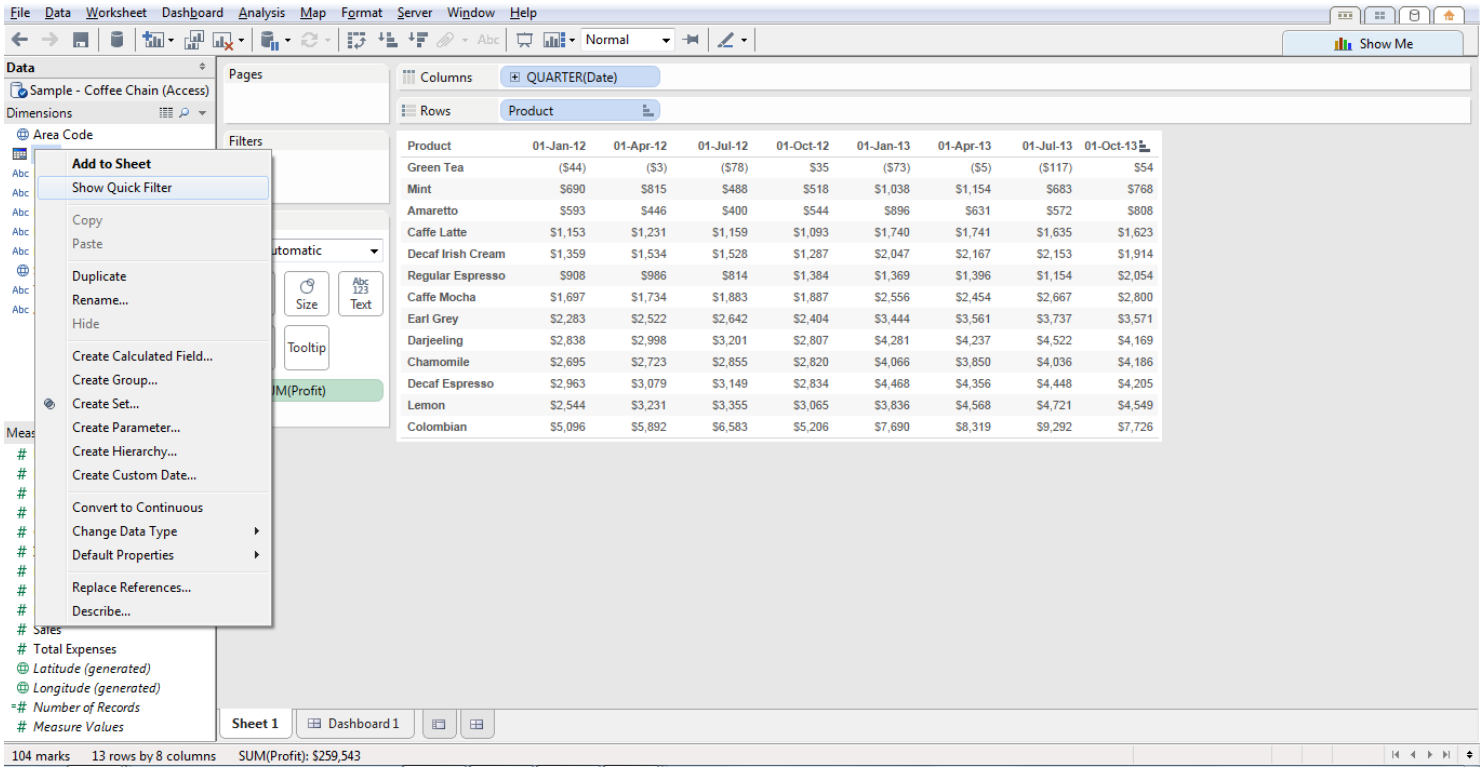
To have a range filter in my dashboard to filter crosstab data, which consist of sales values aggregated by months.

Steps

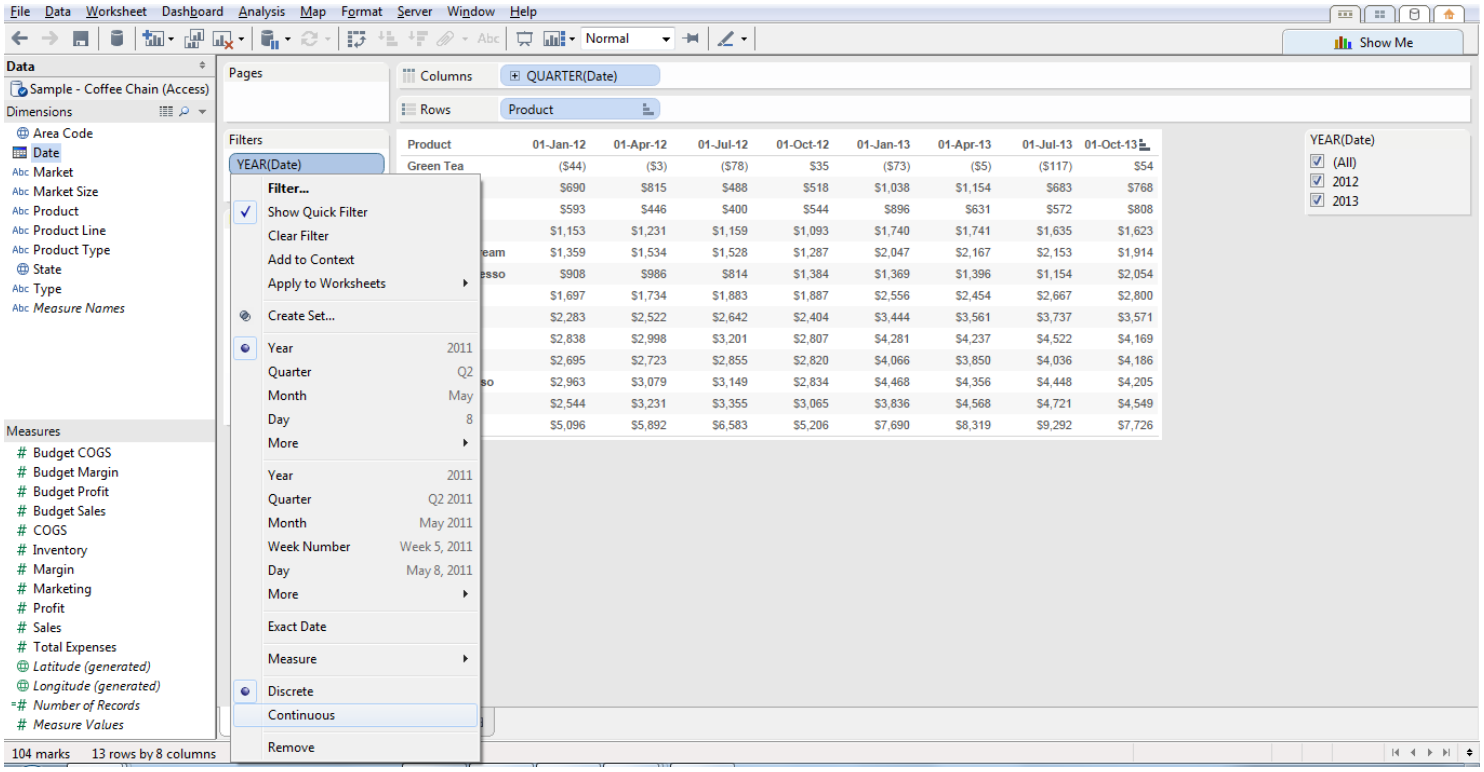
Create a dashboard with Date field in it



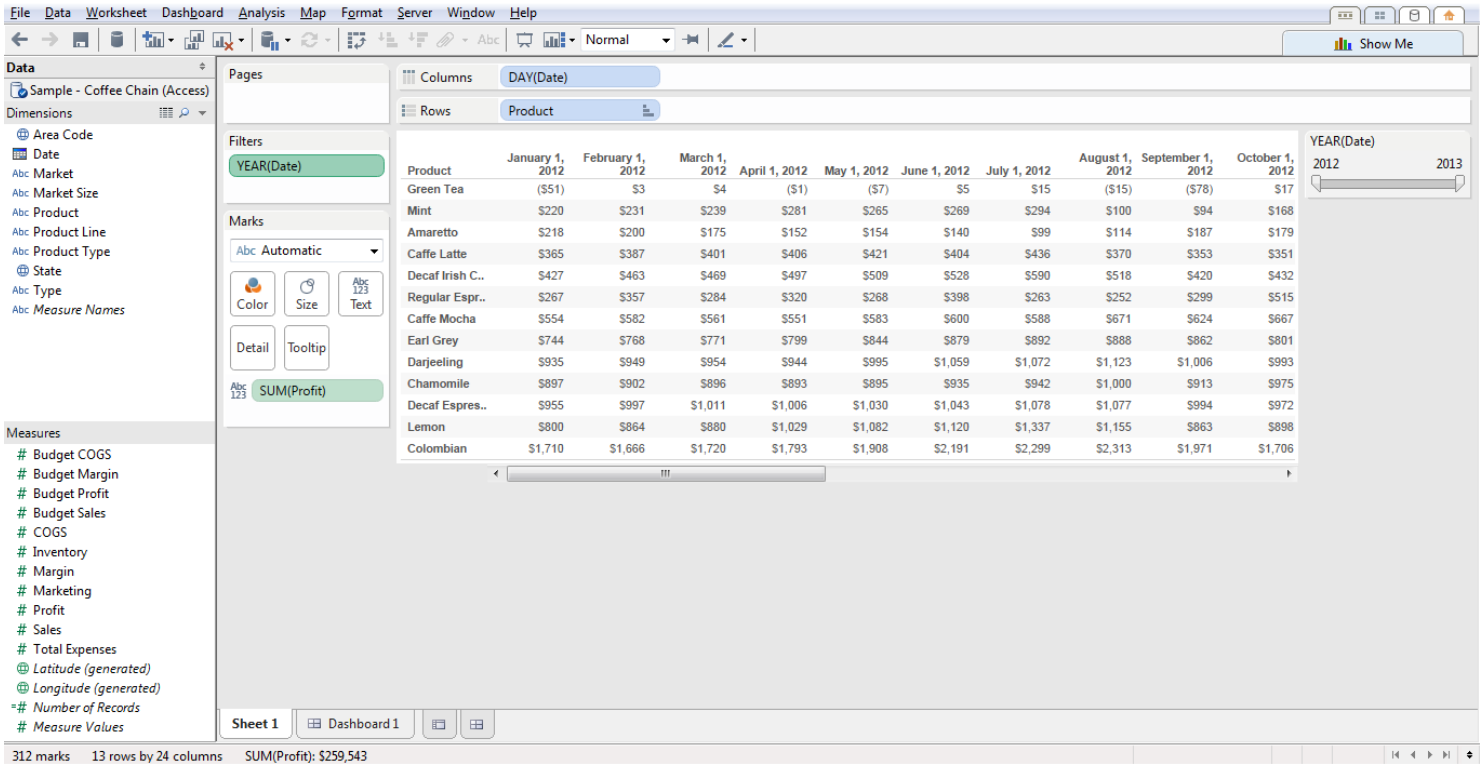
Right-click on the Date field and select “show quick filter”



In the Filter option of the Date field, select “Continuous” instead of “Discrete”

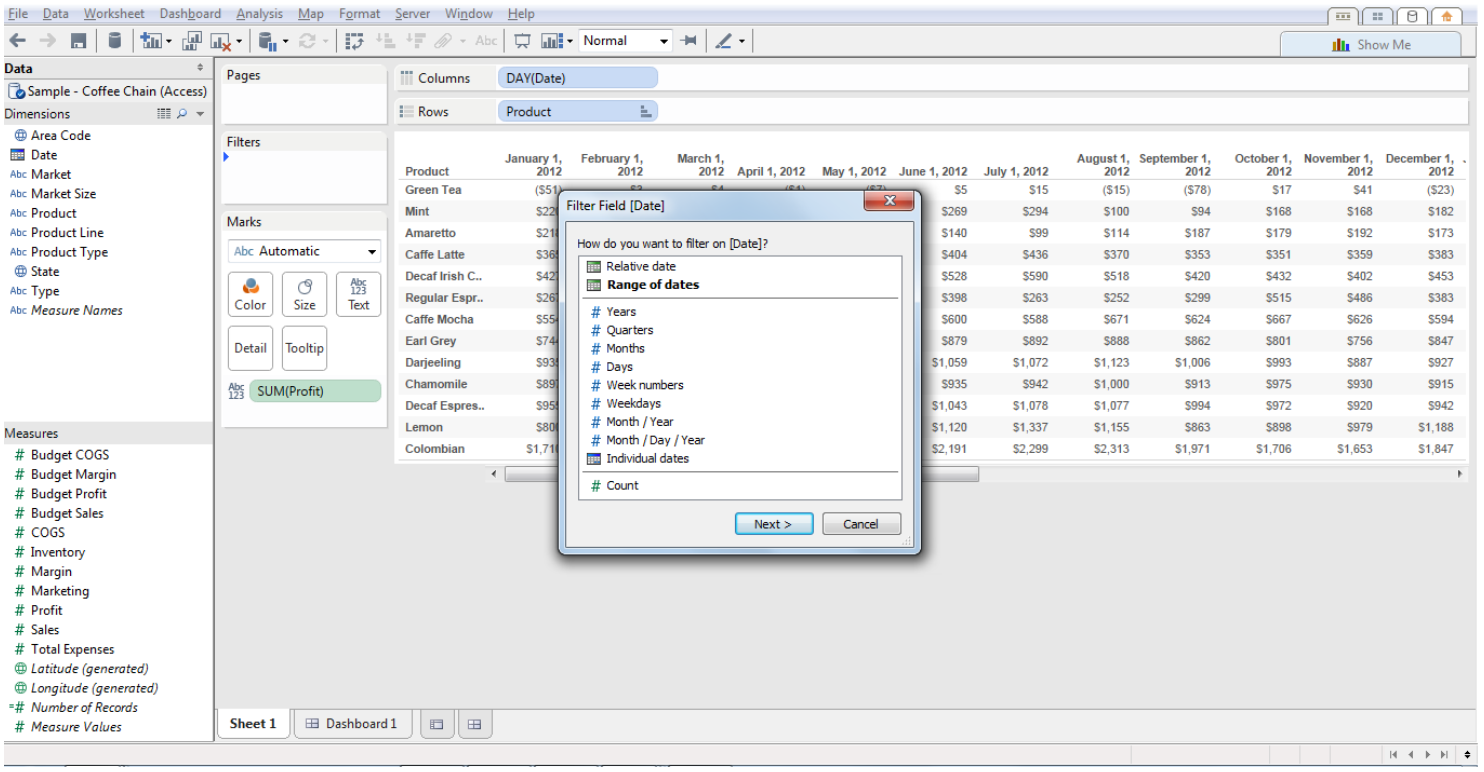


Result



OR

Drag and drop the Date field to the filters mark and you will get all the option related to Date filter



Other Experience Samples

Big Data AchieveMint

By Arunkumar Navaneethan

I had an opportunity to talk to the CEO of AchieveMint, Mikki Nasch about big data and data analytics approaches. The discussions were inclined more towards the healthcare industry. It was definitely a great learning experience.

Before I start, some background about Achievemint. It is company which provides reward points for tracking personal health activities which is already tracked in application. These applications various from wearable fitness monitors like fitbit to social media postings like in Tweeter, foursquare etc.

The company uses this information to understand various patterns and trends with relation to the healthcare industry and the user in specific.

In this blog I will try to highlight the various learning's I had through this conversation.

1. **Business problem definition** must always be the first step. It would be effective to do all the data processing after we know what we are looking for from the data we have.
2. **Usefulness of data**, when data is analyzed its fine to have data which doesn't give any signals even that is an essential learning
3. **Interesting data may not always mean returns**, at times a lot of investment is done on the tools to analysis the data which might provide interesting signals but the returns provided by the signals to the business might be less
4. **Raw data is valuable**, it is important to store and manage all the raw data. The raw data is generally processed and the processed data is stored separately and utilized. They are possibility that certain parts of the raw data were not important during the previous processing, which might be useful for us in the future. Hence, it is essential to store this raw data.
5. **Unsupervised learning model**, unsupervised learning model will not work effectively. It is better to go with the heuristic approach where the domain knowledge drives the analysis to the right and effective direction.
6. **R&D and real-world coexistence**, it is important that R&D and the real world coexist. The results from R&D must be verified time to time with the real world happenings(social media data)

This conversation definitely helped me to understand the real world difficulties of implementing of big data and the challenges an organization would face to gain returns.

Big Data in fraud detection and online marketing – Phoenix Meetup

By Arunkumar Navaneethan

Presenter: Raz Yalov, CTO, 41st parameter

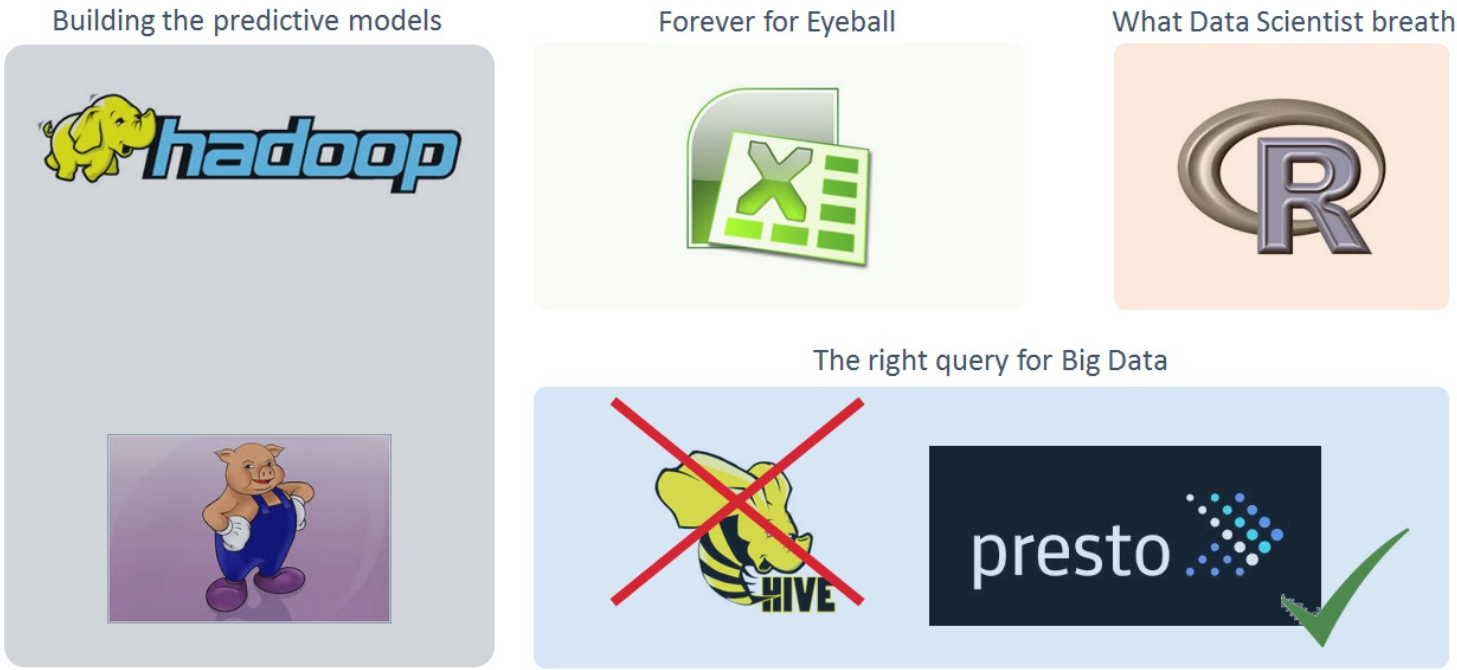
This was one of the best, simple and clear presentation on Big data I have attended.

The presenter distinguishes big data clearly as a challenge of data of size greater than petabytes and data, which we cannot eyeball. We need complex infrastructures only when the regular server cannot handle the processing complexity.

It was interesting to know the difficulties most of the big data analytics companies have with data privacy concerns. The company goes to extend of giving special offers to customers who can provide access to their logs or even real-time sample data to improve the predictive abilities of the analytics system.

Once the customers accepted to sharing the data, the company receives the data using API's, which can be stored into two different storage sources. In-house Hadoop clustered system or Amazons S3 system. The recommended compression formats was .lzo + .lzo.index.

The presenter verdict on various big data technology. The company uses technologies like Pig, Hadoop map reduce, Presto, excel and R programming. Pig and map reduce is used mostly to create the predictive models and excel has been the best resource for eyeballing purposes. The team consist of data scientist who loves R programming, using it as much as possible even if it is not built for that purpose. There was definitely a big dislike towards Hive for its performance issues and connectivity management abilities.



The biggest challenging with Big Data is getting the data itself from the customers and once collected, the next big challenge is to extracting value from it. The sampling processes is one of the major challenges since the level of garbage data present is high.

Looking forward to see how visualization can be injected to such a complex challenging environment.

Big learning from the INSITE big data symposium 2013

By Arunkumar Navaneethan

Conducted by INSITE, Business Intelligence & Analytics Center, University of Arizona

October 10, 2013

Big Data symposium an event organized in University of Arizona, MIS Department, and INSITE center. It had a great list of speakers from diverse backgrounds representing different industrial backgrounds. The symposium covered a huge group of big data use cases from health care to sports.

Started with a great speech from Brian Gentile, Chairman and CEO, jaspersoft, he gave a simple and a clear overview of how big data would influence the industry and discussed the various topics around big data. His speech was about the “New factors of production and rise of big data”. Explained the abundance of data and the value big data brings in by analyzing the data. It was interesting to learn that how big data had been a transitional technical revolution. The transition of the economy from the traditional pattern of land, labor and capital to time, information and capital was an interesting topic. The impact of the four V’s Volume, Velocity, Varity and the fourth new V that I learnt was Varsity. It is essential to understand the importance of the varsity of the analytics achieved from the big data. If the truthfulness of the data is lost, there is no use of the other V’s.
“Applications, Analytics and Aspirations” Brenda Dietrich, VP Strategy and CTO for Business Analytics, IBM. Brenda gave us insights extracting values from the big data analysis. IBM added another V to the stack Value. It is important for industry to understand the value derived before implementing a big data implementation. Big data is not just about social media analytics it is more about extracting value from enormous data available in which we are interested. It is important to understand the internal data before exploring external data sources like the web to conduct your business well. It is safe to start within a controlled boundary where results varied by real experience. With the knowledge gained from that, efforts taken to expand the data sources. It is important that big data analytics provides business value to the organization.
“Driving your business through Big Data”, Darren Stool and Kerem Tomak, marketing analytics, macys.com. It is interesting to see how customer analytics have grown. Importance have changed from understanding the customers implicit feedbacks and shopping preference to the explicit values like how people tend to correlate two different products and similar analysis. Retail industry has been using customer information effectively and the interest appears been growing since the last five years. Macys has been improving the information infrastructure to capture various customer and inventory data to focus and serve its customers better. The retail industry has been focusing on insights to understand its customers better. The industry has also been focusing on using data science and automation in analytics, which appears to be a great indication of big data and analytics growth.

“Unlocking business value with big data and predictive analytics”, Tim Hood, Global VP Chief Solutions Architect, Retail industry, SAP AG. This session explained the importance of big data to carry out predictive analytics. The business value big data could bring through predictive analytics was enormous. Retail industry is one segment where it could be safely and effectively be tested. SAP had taken clear advantage of this using its product HANA. It was interesting to see how predictive analytics and big data was able to function successfully in the gaming world. The ability to use the gaming logs to create retail opportunity is next level of using predictive analytics. It was interesting to see how HANA functions and how it is able to provide real-time analytics from the big data which helps in realizing the business value of data in short interval of time.

Below you can find my real-time opinion as I was part of the symposium through Twitter. Looking forward for the next

years symposium. Will the focus move from Big data to a new concept before 2014? Looking forward for that advancement in analytics.

Business Intelligence and Analytics internship

By Arunkumar Navaneethan

I did my summer internship in Business Intelligence and analytics. It was a career defining experience. Met some interesting and motivating people. Worked very close to senior technology management resources in the company. Every interaction had led to great learning.

Worked in innovation management projects.

The goal was to support innovation within the company seamlessly and also in a simple straightforward fashion. Worked in two projects to support this goal. Created a strawman proposal and a working prototype for data virtualization system. The system will integrate various data sources to form a single point system to search for internal events.

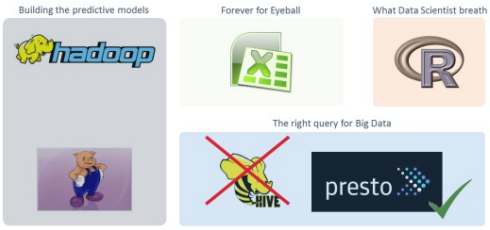
In the second project was able to setup an innovation management system according to business needs. Used a research approach in defining the ideation life cycle process. Considered all the pro's and con's associated with the various ideation approaches and mapped it to the business requirements. Conducted various sprints with the business users to create the system. Along the development process identified various enterprise technical issues and worked with the corresponding teams to get it resolved.

This internship helped me grow professional and come out of the shell of what a technical person must know or do within an organization. It is important to understand the companies mission and align our actions towards the business goals.

Machine data analytics, Big Data meetup

By Arunkumar Navaneethan

Machine data analytics, Big Data meetup



Phoenix, Arizona
Feb 26th 2014

An event conducted by IBM for developers. Presented about machine data analytics and their big data product Big Insight.

Big insight is more of the customized Hadoop system to fit into an enterprise environment. It has various interfaces in it making it an enterprise ready system for big data challenges.

IBM has configured the Hadoop system to be more efficient by utilizing IBM's standard approach like replacing the event pooling system with message pooling system. Using adaptive MapReduce approach using UDP instead of the TCP messaging.

IBM utilizes various accelerators, which is an interface over its data sources like data warehouse, stream and Hadoop systems.

These accelerators helps in creating better insights, normalizes the data, indexes the data for efficient usage by other interfacing systems like visualization suites.

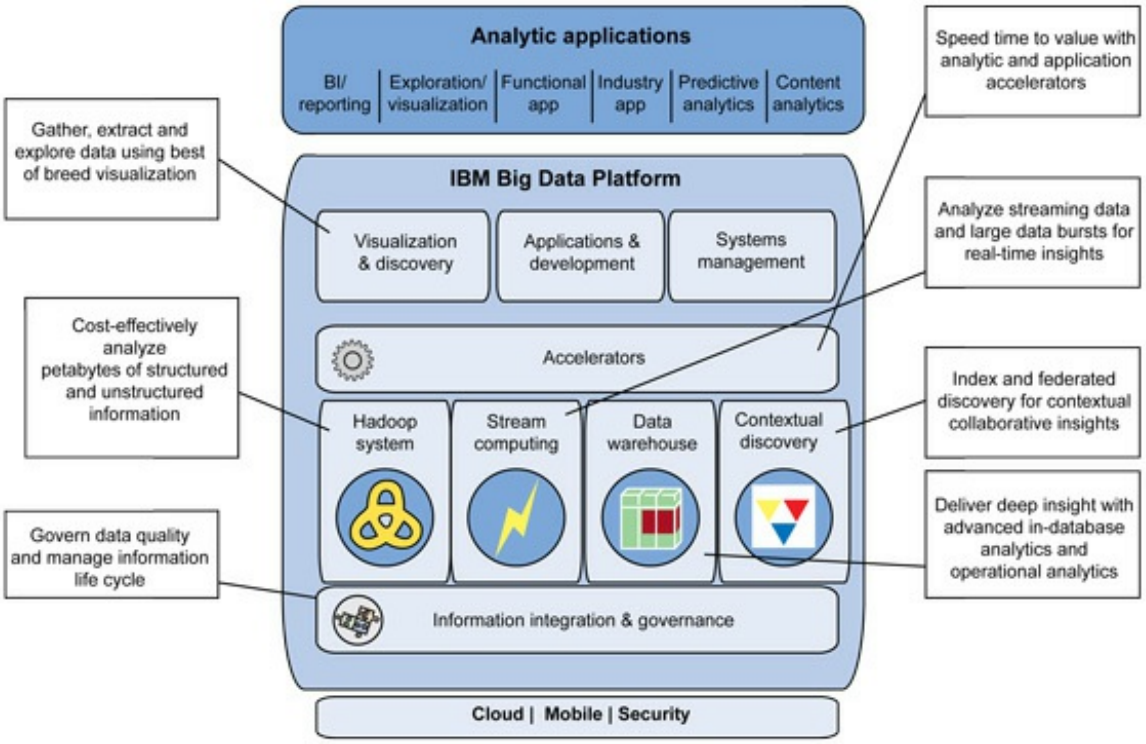
One of these accelerators is the Machine Data Adapter (MDA). The machine data adapter configured with templates of various standard log file structures. Out of the box, it can extract content from log files like web server log, web log and it as capabilities to create and modify these templates for the given machine data.

These templates are the configuration files, the templates can be in different file formats. Other inbuilt data source types are

- Apache Webaccess
- Delimited Separated Values, or CSV files
- Data Power®
- Generic (used when your data source type is not represented by one of the other log types)
- Hadoop Data Node
- Hadoop Jobtracker
- Hadoop Name Node
- Hadoop Secondary Name Node
- Hadoop Task Attempt
- Hadoop Task Tracker
- Syslog, or system log files
- WebSphere® Application Server

It was interesting to see how big data an ambiguous approach is now getting into a more defined enterprise solution suite. It is evident that big data is more about data in rest being utilized effective that just exploring unstructured data. This use case also demonstrates the importance of semi-structure data. When data is semi-structured, the possibility of extracting values from it is high.

I am looking forward for upcoming sessions on AQL, stream computing (real-time data analytics) sessions.

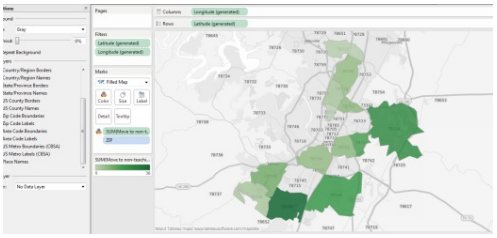


It is interesting that prediction made from data at rest used in real-time using stream analytics approach.

All the observation above, based on my understanding.

Social media analytics – Online news industry

By Arunkumar Navaneethan



Social media has become a major platform that business use to evaluate customer relationships.

For this project, we used Twitter feeds to understand how online news industry is function.

The goal for this project was to find ways by which the online news industry can improve their revenue.

We used a balance scorecard approach to support the goal. Conducted analysis concentrating on perspectives like finance, customer, internal process, and learning and growth opportunities.

The analysis pointed out that customers twitter communications containing reference to the news agency could be helpful.

Customers interaction can be used to understand which kind of articles are been well conceived. Carried out analysis taking into account various interesting parameters, which twitter records.

Article trends and performance helped us understand how events could influence users. The articles temporal information was limited but we were able to use language properties to classify users.

There was presence of group of articles and user by the category of the article. The users were interested in certain types of articles, which drilled down even to the topic.

The devices the users used also affected what there interest was. The devices could help us classify users and there interest on the type of articles. The device level analysis opened up opportunities, which the companies can capitalize.

The propagation of the articles from online sources, social media platforms and twitter users helped us some interesting patterns.

It was interesting to see how our social media behaviors could be used for making business decisions.

Web the laboratory of human impulse

By Arunkumar Navaneethan

(A topic inspired from a lecture by Professor Sudharam, MIS 587; Business Intelligence)

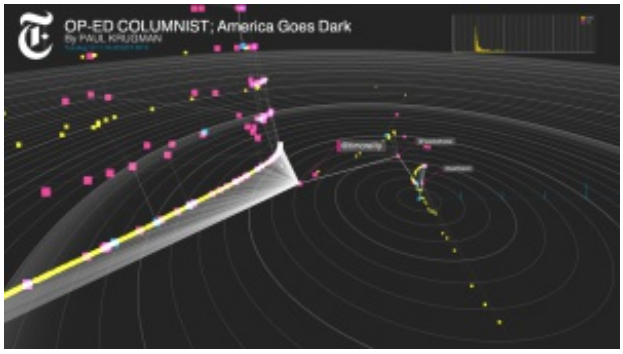
With online activities becoming an essential part of our day-to-day activity, we are leaving behind a lot of information about us in the web. Different companies have been leveraging the data available in the web, majorly the data present in social media sites. This article is about an news agency and how they leverage the web.

News agency New York Times have been doing research on Tweets. It uses an analysis tool called Cascade which analysis the pattern in which New York Times online articles are shared on the web.

Cascade creates a structure which depicts how the published articles propagates in the social media space. This gives them a multi-dimensional view of how the published articles as been conceived by the public. This would be a great platform for them to understand their audience. This gives the news agency the opportunity to understand how long a story published lives in the social media space.

The New York Times
Expect the World®

The tool captures the activities like retweets, replies to the tweets posted by the editors or writers. It also captures tweets which contains the URL to articles published in the New Yorks times website. The tool can capture more interesting data as public share the articles along with their opinion in it.This tool would be available as a Software as a Service, part of the New York Times enterprise initiative; which would capture enterprise bit.ly links shared in through twitter and help organisation capture their social presence in twitter.



References:
http://nytlabs.com/projects/cascade.html

With large data comes great responsilibilites

By Arunkumar Navaneethan

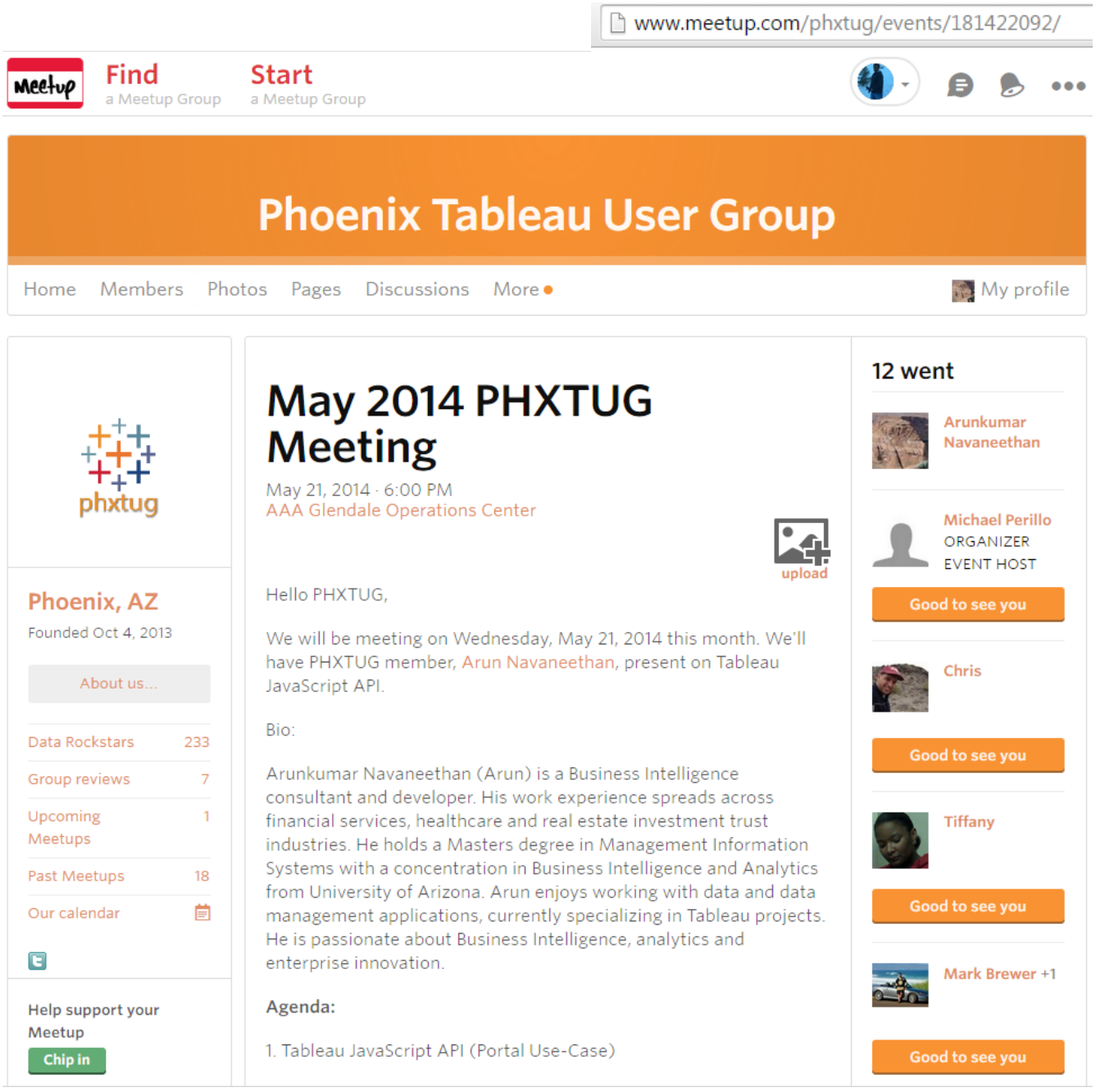
Working with data as always fascinated me. I worked in a financial services company first and was part of the transaction process group. It was amazing how each transaction data was process by multiple teams and the level of value the data had. Each team would process the same data in different ways and the value the data had for every group was different. Understood the various data processing carried out on the data in much better way. Why the data is cleansed, formatted, how the integrated of the data mattered a lot, how a mistake in processing the data can influence the business. Enjoyed this experience of learning data processing in a real corporate environment. Worked in a wide range of database technologies like DB2, Oracle DB and MS SQL. The challenge of understanding a data driven application was very interesting. The value of creating a conceptual schema, relational modeling and designing a relational database with all the integrity constraints made sense.

Now as part of my master's course work, have stepped into a much challenging atmosphere; managing the un-structural data. The days where knowing your data structure before modeling the environment which manages it as slowly being changed. The importance of leveraging textual data like the conversation between people in social media space as grown. Therefore, I am in an interesting place of working with such data. Worked on web and text mining projects and it was quite a challenge. Used Application Programming Interface (API) to extract data from the web related to cities in US. Processed and cleansed this data. Used various data mining algorithms like similarity analysis, clustering and sentiment analysis. Utilized visualization APIs to represent the results in an easily accessible web application. It was a nice learning experience, but now working on much bigger challenge. When we want to extract value from this enormous web, we will need powerful computing abilities. Therefore, the industries buzz word "Big Data" came into picture. My journey with big data has begun where working on analyzing about online tech and social news agencies. Hope my journey with data takes me to better places.

Let data make the world a better place to live in.

Recognitions

Presenter for the Tableau Phoenix User Group – May 2014 on Tableau as a web application



Recommendations Received (1)

Arunkumar, would you like to recommend Arunkumar?
Recommend Arunkumar ▸

Applications Developer

iNautix Technologies, Bank of New York Mellon

Naveen Jesudas
Vice President at iNautix Technologies, a Bank of New York Mellon company

"Arunkumar Navaneethan was self-motivated and a self-learner in my team. He was able to learn the IBM Lombardi Teamworks BPM tool in a very short period of time on his own. He led the vendor transition efforts in taking over the production support activities of the WebSTP BPM application. He was also able to train the other team members on Lombardi. This was invaluable since recruiting Lombardi resources from the market can be quite challenging.

Arunkumar showed strong initiative in organizing activities for the corporate social responsibilities events within iNautix. He was also an outgoing and invaluable member of my team. **less**

January 17, 2013, Naveen managed Arunkumar at iNautix Technologies, Bank of New York Mellon