Team Unknown - CS235

Milan Potdar (milan.potdar@sjsu.edu)
Huy Vu (huyvu262@yahoo.com)
Nachappa AP (nachappa.ap@sjsu.edu)

Data Visualizations Project Documentation
California Education Data

If a picture is worth a thousand words, a good data visualization must always be better than a table.

1. **Introduction**
2. **Purpose**
3. **Target Users And Goals**
4. **Data Source**
5. **Technologies Used**
6. **Design Patterns**
   a. **Navigation Model: hub and Spoke & Clear Entry Point**
   b. **Data Visualization Patterns**
7. **Justification for Visualizations**
   Map
   Bubble Chart
   Lien Chart
   Cycles
   Exceptions
   Dot Plot
   Heat Maps
   Pie Charts
   Data Ink Ratio
   Bar Charts
8. **Data Visualization**
   Enrollment by Country
   Test Popularity
   Performance by County
   CST English performance High School
   CST Math Performance High School
   High school in Santa Clara County
   Elementary school in Santa Clara County
   Trend Dashboard
   Grad Percentage vs Ethnicity
   Ethnicity vs UC Grad
   Grad VS undergrad and Ethnicity
   Grad percentage vs Under grad percentage
   Ethnicity vs suspension
   Suspension type vs reasons
   Suspension types vs reason vs year
   Enrollment vs Drop Out
   Language count in CA
   Native Languages speaker in District school
   Poverty and School Education
   Transportation mode
   Sports in school
9. **How to Run Application.**

INTRODUCTION

Data use in increasing rapidly within the education space. Policy makers are under increasing pressure to use data to inform decisions, justifying funding and guide practice. Moreover parents who are looking to get there children admitted in the school are also get benefitted by this data. And schools also need an effective way to check on its own performance.

Thus we see that School, Law Makers and Parents are the people who will get benefited from this visualization. Schools, Colleges and Universities that receive funding are required to submit data annually which means you can find information on almost any of these institutions of our liking. Unfortunately, the data is not adjudicated. The surveys are completed by people who don't understand the business or the system they are being surveyed about. Thus may not be reliable and are time consuming process.

As policy makers are often not statisticians, researchers or quants. Thus Data visualizations is a way to bridge this gap. Proper data visualization will bring the data to the audience in a way they can understand quickly and use to inform decisions.

PURPOSE

Why to visualize the data?

The purpose of graphical display is to provide the viewer a visual means of processing the information. It is important to note visualization to be effective it must draw upon the knowledge of the viewer. Thus it helps us in data Exploration and best way to communicate information. Data visualization is a tool for communicating a specific feature of a dataset in an approachable and efficient manner.

Thus visualizing the data makes it easy for people who are form this domain to draw conclusions they are looking for without any hard work and just by hovering or on simple click.

TARGET USERS AND GOALS

Who will use it and what are their goals?

Data visualization on education data is of great help for Schools, Law Makers and Parents.

- Schools need data to improve, but balancing the data, analysis and reporting requirements across a system of schools can be challenging.
- Data visualization can effectively monitor and analyze student test performance.
- Understand and share enrollment statistics in minutes.
- Easily evaluate the performance of schools and teachers.
- Rapidly share insights across your organization.
- Evaluate student performance against institutional goals.
- Analyze student data: enrollment statistics, achievement, and demographics.
- Parents can see and filter schools as per there requirements.
- Law makers can analysis and see if Law's are followed.

DATA SOURCE

The data for this visualization project is taken from California Education data. It's a public data set collected by California government by various methods and is published every year. It consists of many data set with variety of data. All the information can be found on www.cde.ca.gov .

It has total of 16 csv files. For our analysis we took data from 2008 to 2014. 2015 data is not yet released yet.

TECHNOLOGY USED

Splunk for data analysis , Excel , Tableau , HighCharts ,BlogSpot

Design Pattern

Navigation Models
Hub and Spoke pyramid as everything is on the main page of the blogger
Clear entry point as it starts from the main link of the blog with is introduction of the blog.

Analytic interaction with data:
Comparing: Ranking, Part to whole, Time series, scatter plot
Sorting, Filtering, highlights, Re-Scaling, Annotating, and accessing details on demand.

Data visualization patterns:
Data Tips, Data Spotlight, Dynamic Queries, multivariate displays: Glyphs

JUSTIFICATION FOR VISULIZATION

Map: These days with the advent of google maps, everyone is inclined to use maps when location/ demographic details are in picture. Thus in order to distinguish based on demographic we used maps. Here we distinguish based on County thus best thing to do was to show it on maps.

Bubble Charts: These are used when we are trying to compare based on amounts and quantity. It clearly shows which is dominant in the group.

Line Chart: used for following reasons:

> **Trend:** the overall tendency of series of values to increase, decrease, or remain relatively stable   during particular period of time.
> **Variability:** the average degree of change from one point in time to next during particular period of time.
> **Comparing** two set of data
> **Rate of change:** when change is in percentage difference.
> **Co-variation:** two time series relate to each other even moving in opposite direction.
> **Cycles:** Patterns that repeat at regular time intervals.
> **Exceptions:** Values that fall outside the norm.

Dot Plot : When difference in qualitative analysis is shown.

Heat maps: used in maps to show data.

Pie Chart: pie charts are used only when there is prominent difference in two data groups related to each other thus making area of pie slices comparable easily. Mostly replace by Bar charts.

Data Ink ratio: every bit of Ink present new information from same domain and thus we can make comparison very easily.
Bar charts: Highlight low level detail item that need viewer attention. Make comparisons within data values. Compare and contrast multiple sets of values.

DATA VISULIZATIONS

Description: This map shows the number of students enrolled in STAR within each county. Hovering on each region on the map shows the enrollment information of that county based on districts, schools, and students, total STAR enrollment.

Visualizations: Map, Data Tips, Data Spotlight, Dynamic Queries.

Conclusion: Los Angeles has highest enrollment in California followed by San Bernardino, River side and San Diego.

Description: CST Math and English are the most popular test, since the subject are the requirement of K-12 education. CST is the standard test to examine students' skills. Because English and Math are the most popular, it is hard to see how many times other tests are taken, but people can still view this by selecting the test name on the radio panel on the right. Hovering on each region shows Test Name and total people who gave the test.

Visualizations Pattern: Bubble Chart, Data Tips, Data Spotlight, Dynamic Queries.
Conclusion: Mathematics and English Language are taken the most. It is also because it's mandatory. Thus law makers can easily make out from this graph

School Performance

Description: this chart shows student competency composition of each county, allowing viewer to compare performance of schools in each county. Each bar shows how students of its associated county are divided based on their level of competency from advanced to far below average. Viewers can move position of the sub-bar for easier comparison between counties.

Visualizations Pattern: Data Ink Ratio, Quick Data Filters, Dynamic Query, Spot light, Data tip

Conclusion: Can be used to make prediction of the performance based on the filters and see where students have above average performance.

School Performance
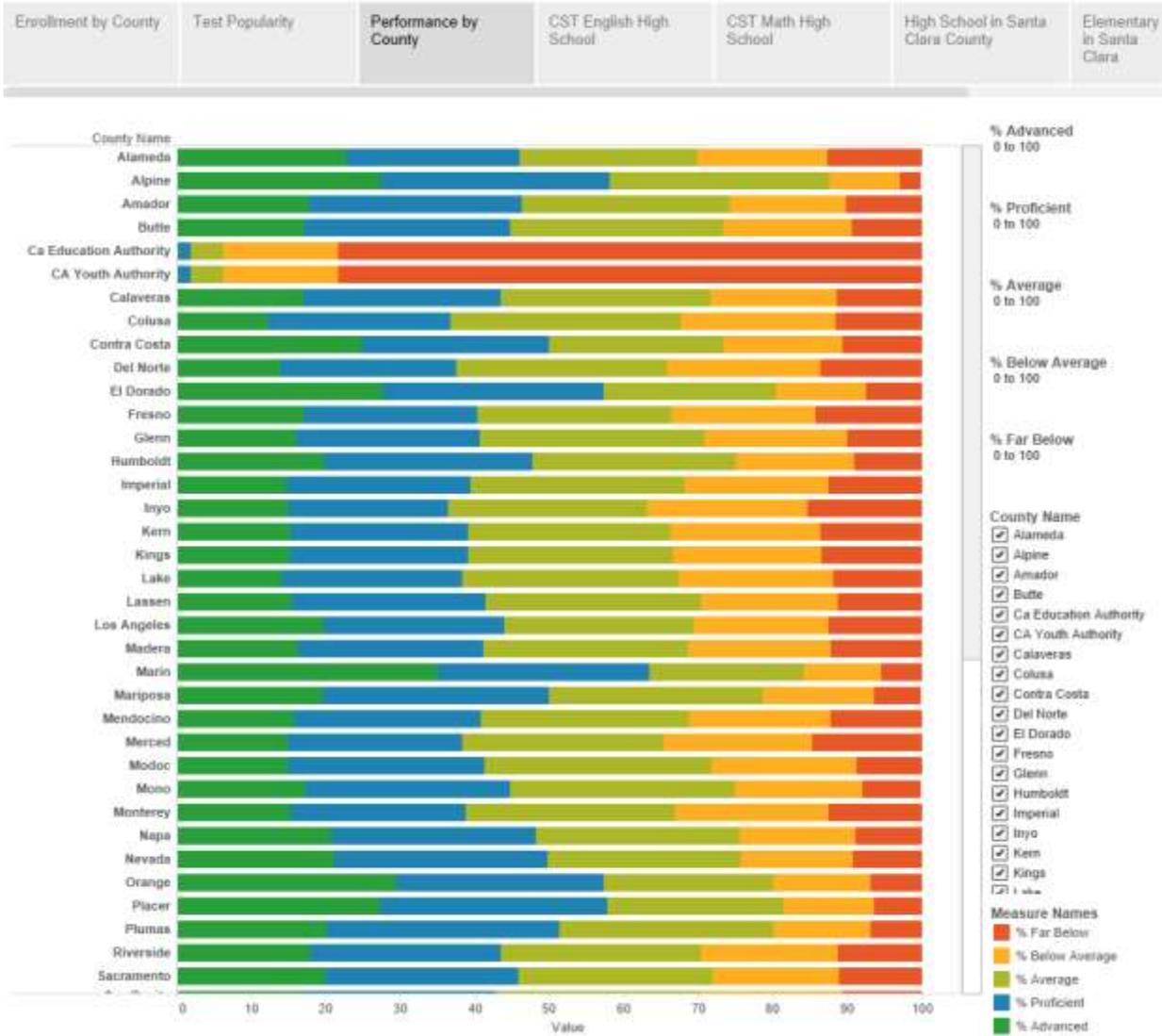
Description: statistics of CST English for high school. The dashboard shows statistics of CST English performance by high school (this level of education is more concerned than elementary and intermediate). The filters on the right panel allow viewers to perform dynamic queries for easier comparison.

Visualizations Pattern: Bar Charts and Tables on Dashboard, Data filters, Text Based search, Tool tip and spot light.

CONCLUSION: Shows Stats of CST English Performance on user based location using Zip Code and county names and thus also show average mean score for students.

School Performance

CST English Chart

Avg. Mean Scale Score
224.8 ──── 454.1

Test Name
CST English-Language Arts

County Name	District Name	School Name
Alameda	Alameda City Unified	Alameda Comm Learning
		Alameda Community Learning
		Alameda High
		Alameda Science & Technolo.
		Alameda Science and Techno.
		Encinal High
		Island High (Continuation)
	Alameda Community Learning.	Alameda Community Learning.
	Alameda County Office of Education	Alameda Co. Opportunity
		County Community
		Juvenile Hall/Court
	Albany City Unified	Albany High
		MacGregor High (Continuation)
	American Indian Public High	American Indian Public High
	ARISE High	ARISE High

Avg. Mean Scale Score

District Name

School Name

Avg. Mean Scale Score
147.7 to 600

Zip Code
- ☐ 64621
- ☑ 65367
- ☑ 65531
- ☑ 84607
- ☑ 89447
- ☑ 90001
- ☑ 90002
- ☑ 90003
- ☑ 90004
- ☑ 90005
- ☑ 90006
- ☑ 90007
- ☑ 90008
- ☑ 90010
- ☑ 90011

County Name
- ☑ Alameda
- ☑ Alpine
- ☑ Amador
- ☑ Butte
- ☑ Ca Education Authority
- ☑ CA Youth Authority
- ☑ Calaveras
- ☑ Colusa
- ☑ Contra Costa
- ☑ Del Norte
- ☑ El Dorado
- ☑ Fresno
- ☑ Glenn
- ☑ Humboldt
- ☑ Imperial

CST English Stat

Test Name / Grade
CST English-Language Arts

County Name	District Name	School Name	9	10	11
Alameda	Alameda City Unified	Alameda Comm Learning	393.7	376.9	389.4
		Alameda Community Learning	393.7	376.9	389.4
		Alameda High	384.7	366.1	364.2
		Alameda Science & Technolo.	485.5	390.8	369.9
		Alameda Science and Techno.	485.5	390.8	369.9
		Encinal High	355.0	341.1	338.3
		Island High (Continuation)		276.2	291.0
	Alameda Community Learning.	Alameda Community Learning.	393.7	376.9	389.4
	Alameda County Office of Education	Alameda Co. Opportunity	301.9	294.4	271.9
		County Community	290.2	278.1	269.7
		Juvenile Hall/Court	277.1	259.3	242.4
	Albany City Unified	Albany High	389.6	375.3	374.9
		MacGregor High (Continuation)		294.3	270.9
	American Indian Public High	American Indian Public High	422.7	420.5	411.0
	ARISE High	ARISE High	311.7	299.5	296.8

Description: statistics of CST Summative Math and Algebra II for high school. The dashboard shows statistics of CST Summative Math and CST Algebra II (2 major math tests for high school students) performance by high school (this level of education is more concerned than elementary and intermediate). The filters on the right panel allow viewers to perform dynamic queries for easier comparison.

Visualizations Pattern: Bar Charts and Tables on Dashboard, Data filters, Text Based search, Tool tip and spot light.

CONCLUSION: Shows Stats of CST Math Performance on user based location using Zip Code and county names and thus also show average mean score for students.

School Performance

| Test Popularity | Performance by County | CST English High School | CST Math High School | High School in Santa Clara County | Elementary in Santa Clara | Trend Dashboard |

CST Math Chart

Avg. Mean Scale Score
191.6 — 532.7
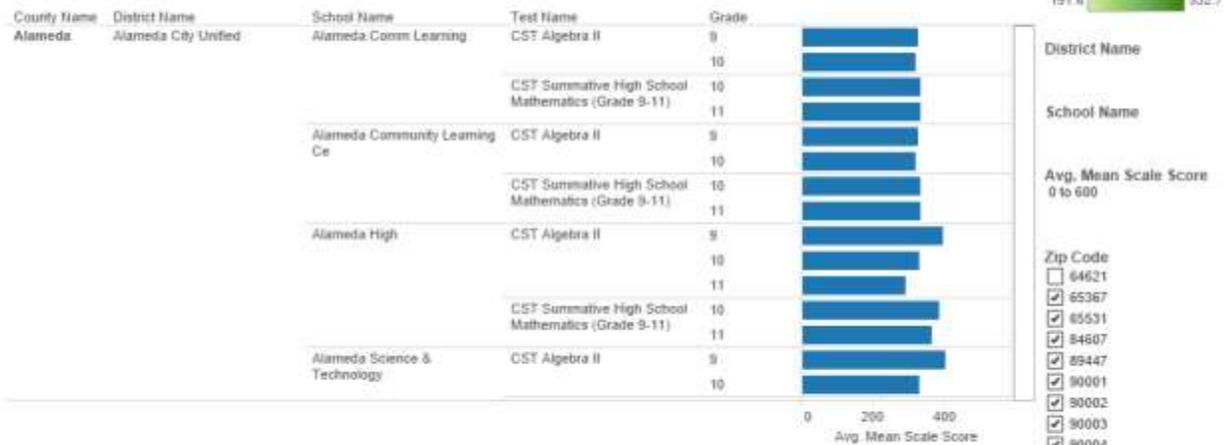
County Name	District Name	School Name	Test Name	Grade
Alameda	Alameda City Unified	Alameda Comm Learning	CST Algebra II	9
				10
			CST Summative High School Mathematics (Grade 9-11)	10
				11
		Alameda Community Learning Ce	CST Algebra II	9
				10
			CST Summative High School Mathematics (Grade 9-11)	10
				11
		Alameda High	CST Algebra II	9
				10
				11
			CST Summative High School Mathematics (Grade 9-11)	10
				11
		Alameda Science & Technology	CST Algebra II	9
				10

Avg. Mean Scale Score (axis: 0, 200, 400)

Right panel filters:

District Name

School Name

Avg. Mean Scale Score
0 to 600

Zip Code
☐ 64621
☑ 65367
☑ 65531
☑ 84607
☑ 89447
☑ 90001
☑ 90002
☑ 90003
☑ 90004
☑ 90005
☑ 90006
☑ 90007
☑ 90008
☑ 90010
☑ 90011

County Name
☑ Alameda
☑ Alpine
☑ Amador
☑ Butte
☑ Ca Education Authority
☑ CA Youth Authority
☑ Calaveras
☑ Colusa
☑ Contra Costa
☑ Del Norte
☑ El Dorado
☑ Fresno
☑ Glenn
☑ Humboldt
☑ Imperial

CST Math Stat

Test Name / Grade

County Name	District Name	School Name	CST Algebra II			CST Summative High School Ma...		
			9	10	11	9	10	11
Alameda	Alameda City Unified	Alameda Comm Learning	328.7	318.8			334.0	335.6
		Alameda Community Learning.	328.7	318.8			334.0	335.6
		Alameda High	395.6	330.2	293.2		385.0	366.4
		Alameda Science & Technolo.	404.4	330.9			392.0	321.5
		Alameda Science and Techno.	404.4	330.9			392.0	321.5
		Encinal High	377.2	324.7	309.0		383.8	338.2
	Alameda Community Learning.	Alameda Community Learning.	328.7	318.8			334.0	335.6
	Albany City Unified	Albany High		410.0	342.6			405.9
	American Indian Public High	American Indian Public High		454.0				434.6
	ARISE High	ARISE High		249.1	248.0			
	Aspire California College Pre	Aspire California College Pre		340.0	311.1			317.3
	Aspire California College Prep	Aspire California College Prep		340.0	311.1			317.3
	Aspire Golden State College Pr	Aspire Golden State College Pr	277.1	266.2				327.0
	Aspire Lionel Wilson College	Aspire Lionel Wilson College	329.4	331.3	274.7			302.3
	Aspire Millsmont Secondary Ac	Aspire Millsmont Secondary Ac	277.1	266.2				327.0

Description: Comparison of high school in Santa Clara County. This charts compares the percentage of proficient and advanced students based on their test score for schools within Santa Clara County. An interesting correlation is that school performance is directly proportional to house value and real estate market activity in area near the schools. Viewers can select to view any set of school for easier comparison.

Visualizations: Data Ink Ratio, text Based search, dynamic axis of Reference line showing minimum proficient percentile, label.

School Performance

Test Popularity	Performance by County	CST English High School	CST Math High School	High School in Santa Clara County	Elementary in Santa Clara	Trend Dashboard

District Name	School Name	% Highend Student	% Average or Below
Campbell Union High	Boynton High	12.3	87.9
	Branham High	40.8	59.2
	Camden Community Day	7.4	92.5
	Del Mar High	23.0	77.0
	Leigh High	46.3	53.7
	Prospect High	45.2	54.9
	Westmont High	39.8	Minimum Proficient Percentile
East Side Union High	Apollo High	7.4	92.7
	Apollo High Cont.	7.4	92.7
	Evergreen Valley High	51.3	48.7
	Foothill High	7.6	92.4
	Genesis High Cont.	13.3	86.5
	Hill (Andrew P.) High	29.0	70.3
	Independence High	35.7	64.3
	Lick (James) High	19.9	80.1
	MACSA Academia	7.5	92.6
	MACSA Academica	7.5	92.6
	Mt. Pleasant High	26.5	73.6
	Oak Grove High	28.6	71.4
	Overfelt (William C.) Hig	20.2	79.9
	Pegasus High Cont.	4.4	95.7
	Phoenix High Cont.	5.3	95.0
	Piedmont Hills High	47.9	52.1
	Santa Teresa High	37.9	62.1
	Silver Creek High	39.7	60.2
	Yerba Buena High	25.8	Minimum Proficient Percentile
Fremont Union High	Cupertino High	58.6	41.4
	Fremont High	32.9	67.2
	Homestead High	55.4	44.6
	Lynbrook High	76.1	24.0
	Monta Vista High	73.8	Minimum Proficient Percentile
Gilroy Unified	Christopher High	41.2	58.9
	Dr. T. J. Owens Gilroy Early	69.3	30.7
	Dr. T. J. Ownes Gilroy Early	69.3	30.7
	Gilroy Community Day	100.0	
	Gilroy High	31.1	68.9

School Name

Measure Names
- % Average or Below
- % Highend Student

Minimum Proficient Perc..
60

Value axis: 0 10 20 30 40 50 60 70 80 90 100

Description: comparison of elementary in Santa Clara County. An alternative perspective at elementary level, showing students perform better at young ages, before they become lazier in school work

Visualizations: Data Ink Ratio, text Based search, dynamic axis of Reference line showing minimum proficient percentile, label.

School Performance

| Test Popularity | Performance by County | CST English High School | CST Math High School | High School in Santa Clara County | **Elementary in Santa Clara** | Trend Dashboard |

School Name

District Name	School Name		
Gilroy Unified	Aprea (Luigi) Elementary	88.90	31.10
	Del Buono (Antonio) Elementa..	49.15	50.83
	El Roble Elementary	50.16	49.88
	Eliot Elementary	54.58	45.31
	Glen View Elementary	47.84	52.26
	Las Animas Elementary	65.48	34.61
	Rod Kelley Elementary	56.15	43.87
	Rucker Elementary	54.49	Minimum Proficient Percentile
Palo Alto Unified	Addison Elementary	89.44	10.64
	Barron Park Elementary	70.64	29.31
	Briones (Juana) Elementary	85.62	14.51
	Duveneck Elementary	93.67	6.20
	El Carmelo Elementary	87.84	12.98
	Escondido Elementary	82.13	17.72
	Fairmeadow Elementary	85.73	14.38
	Hays (Walter) Elementary	86.84	12.96
	Hoover (Herbert) Elem	95.71	4.33
	Nixon (Lucille M.) Elementary	89.40	10.69
	Ohlone Elementary	82.02	17.98
	Palo Verde Elementary	87.56	Minimum Proficient Percentile 12.33
San Jose Unified	Allen at Steinbeck	52.40	47.71
	Almaden Elementary	45.32	54.67
	Bachrodt (Walter L.) Elementar	45.10	54.90
	Booksin Elementary	79.55	20.43
	Canoas Elementary	44.02	56.03
	Carson (Rachel) Elementary	66.84	33.07
	Darling (Anne) Elementary	46.29	53.74
	Empire Gardens Elementary	34.07	65.80
	Galarza (Ernesto) Elementary	48.09	52.08
	Gardner Elementary	49.74	50.23
	Grant Elementary	44.63	55.27
	Graystone Elementary	88.93	11.29
	Hacienda Magnet	70.89	28.87
	Hammer Montessori at Galarza	67.47	32.58
	Los Alamitos Elementary	83.02	16.98
	Lowell Elementary	53.64	46.43

Measure Names
- % Average or Below
- % Highend Student

Minimum Proficient Perc..
60

Value axis: 0 10 20 30 40 50 60 70 80 90 100

Description: Trends of STAR performance in larger counties. This shows trends in 6 largest counties in California, where people tend to find jobs, buy houses, and send their children to school. Viewers can select to highlight a single trend to see important information throughout the year

Visualizations Pattern: Line Charts, Spot light, dynamic data filters based on county names, tool Tips.

School Performance

| Test Popularity | Performance by County | CST English High School | CST Math High School | High School in Santa Clara County | Elementary in Santa Clara | **Trend Dashboard** |

Big County % Test Trend — Test Year — Avg. Percent Tested

Big County MeanScore Trend — Test Year — Avg. Mean Scale Score

Big County %HighEnd Trend — Test Year — Value

County Name
- Los Angeles
- Orange
- Sacramento
- San Diego
- San Francisco
- Santa Clara

Description: This shows grad percentage of various ethnicity and compares it with the gender. It shows grad percentage, name of ethnicity, total number of grad and its percentage.

Visualizations Pattern: Bar Chart, Spot light, Tooltip, dynamic filters, label.

Conclusion: Male vs Female ratio are equal and thus abide by the law.
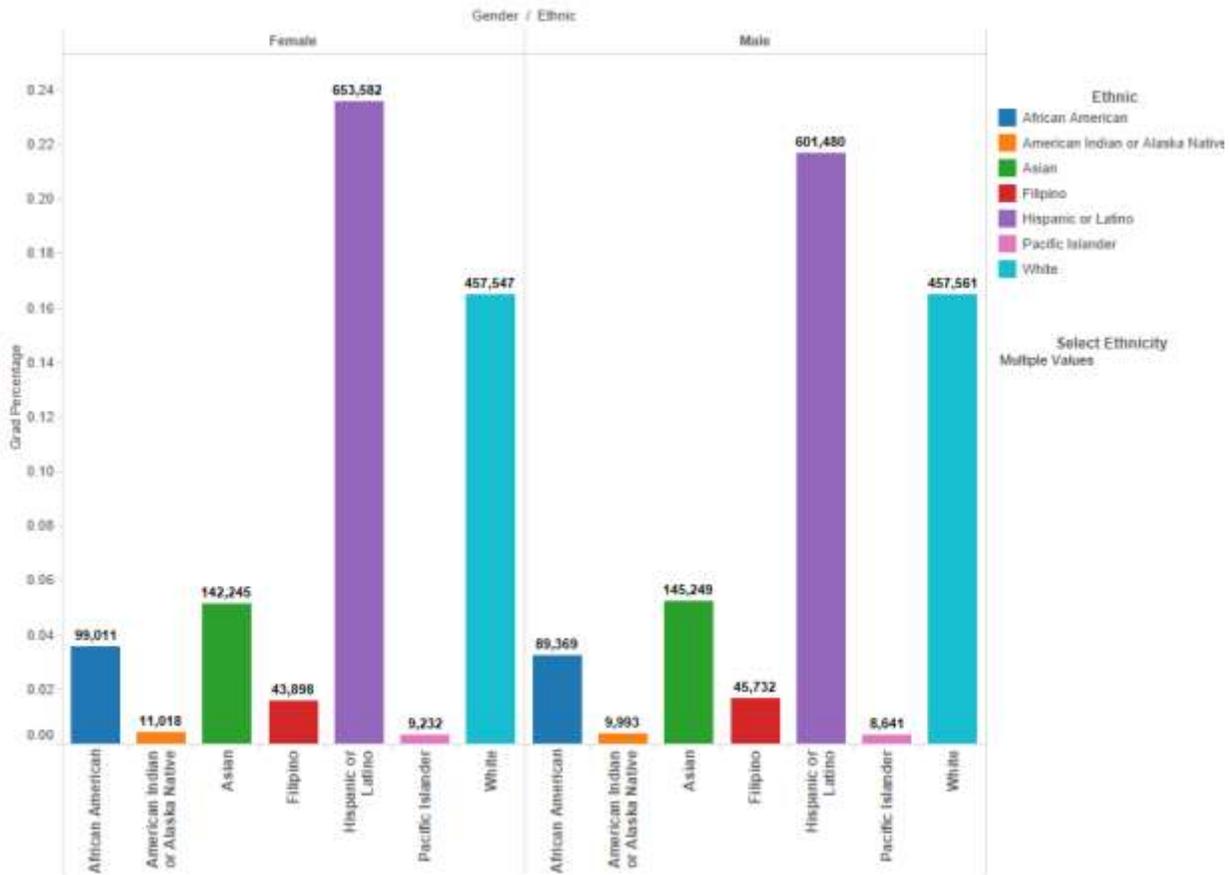
Description: Line chart shows the number of graduates based on ethnicity over the years.
Visualizations: Tool tip, multiline charts, dynamic quick filter, spot light.
Conclusion: We can see that number of Hispanic and Latin people getting admission in UC and at grad level have increased exponentially.
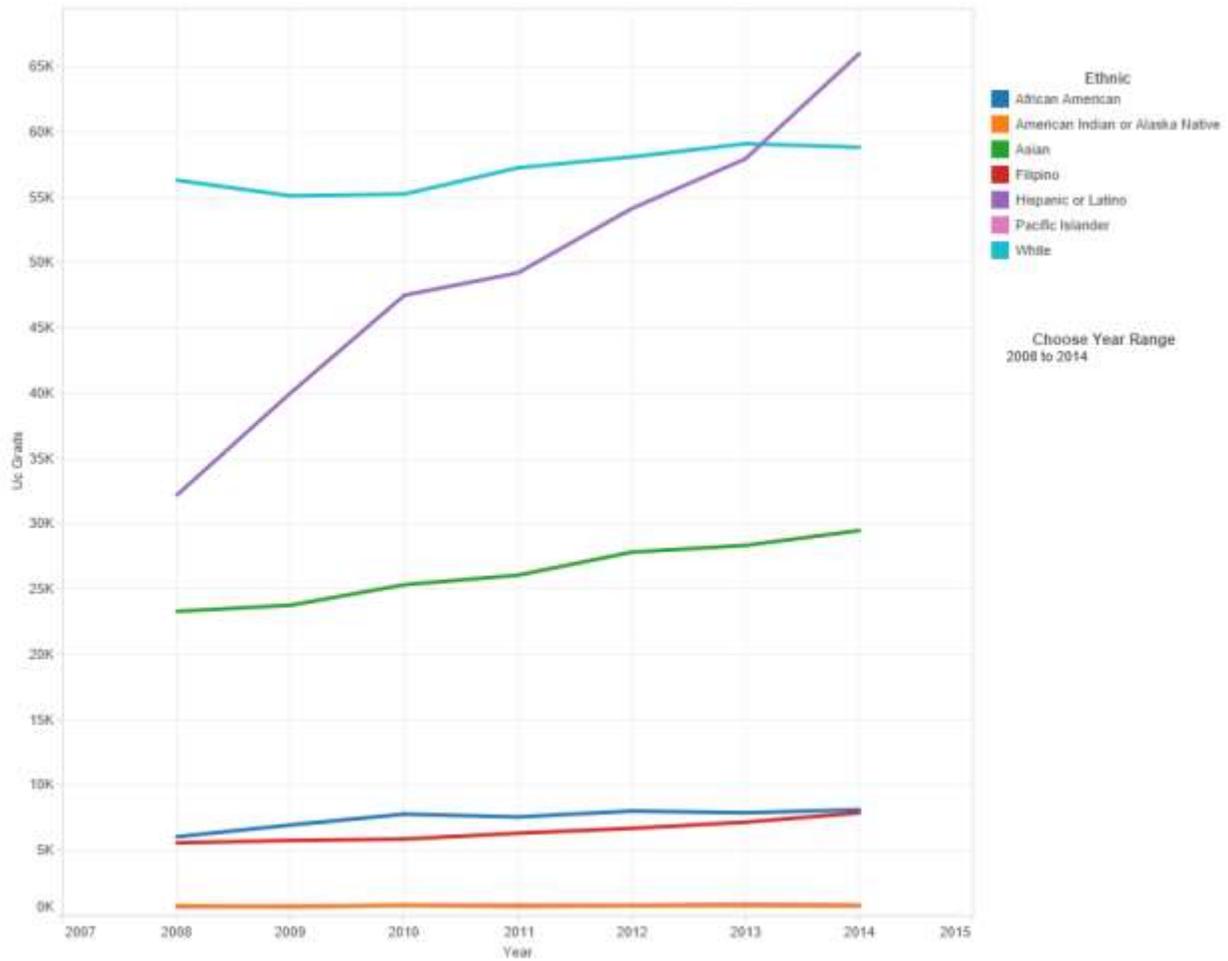
Description: Bar chart shows number of graduating students with respect to gender and ethnicity. There are approximately equal number between male and female students. Meanwhile, some ethnics dominates over other despite California is among the most ethnical diverse states

Visualizations Pattern: Bar Chart

Conclusion: Asian people acquires maximum seat as UC grad.
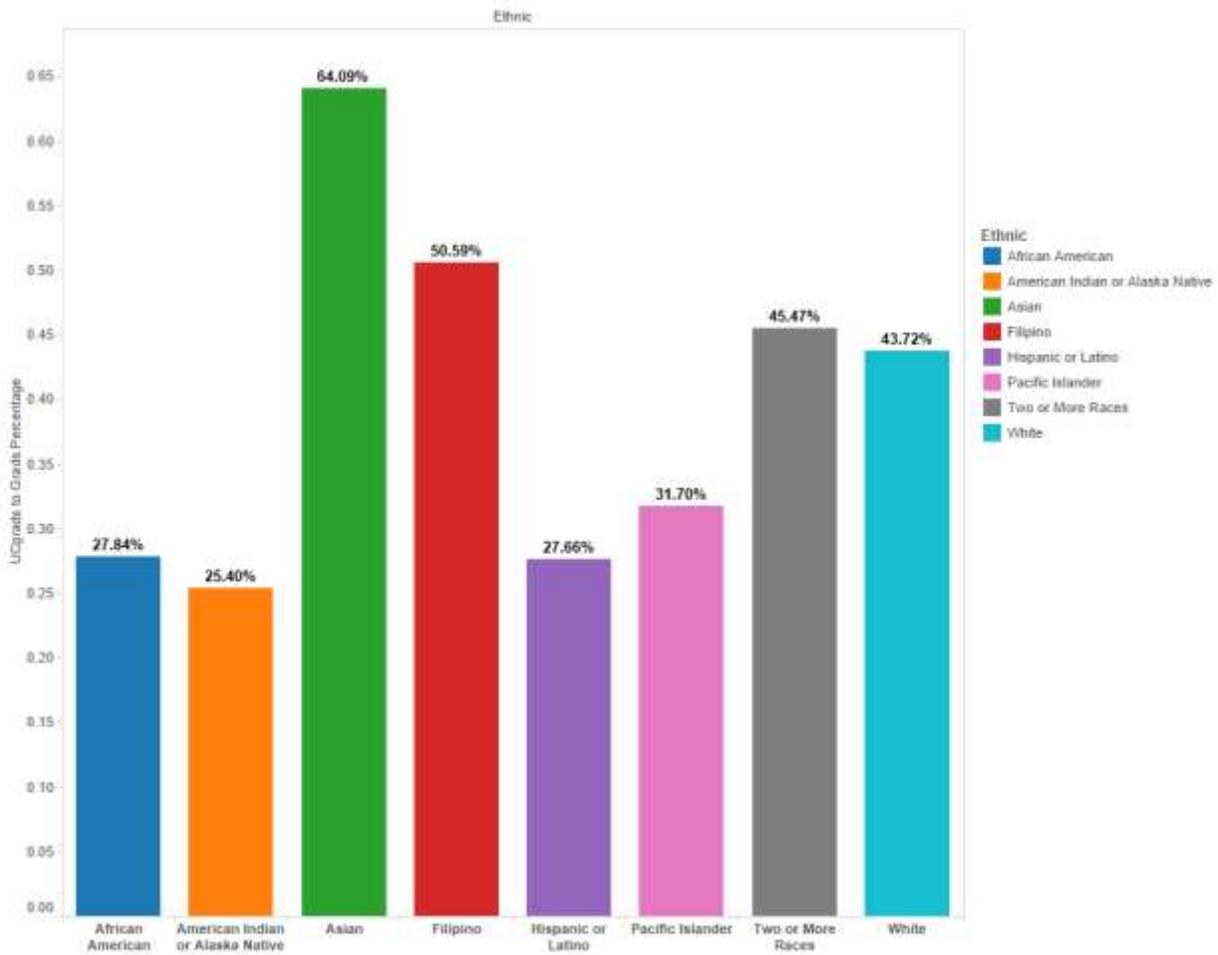
Description: Show the graph of ethnicity vs suspension. It also has a dynamic query when you click on right side on Disciple type. It has three types. Expulsion, In-School suspension and out of school suspension. And also shows reason of suspensions like violence without injury, violence with injury, drugs and weapons.

Visualizations Pattern: Multi Line chart, Spot light, dynamic query on Discipline Type and measure names.

Conclusion: This Graph show s with glaring clarity that Violence committed by Hispanic people is very high but the point to note here is why is the case. Studies show that it's debatable to relate race and crimes. But there are many more factor's that should be taken into consideration while releasing such data into public. This may hurt people's emotion as no other data is released into public except this data set. If it may have shown the background of the student or other factor were taken into consideration it can be easily justified as false figure which does not take into consideration the whole scenarios and displays just the data with no clear picture of real life scenario. The other scenarios may be the total population count, background of the student, number of white people in power, how many were actually found guilty and were not guilty etc. things should also be made as part of this data set. Thus shows us an area where data set should be refined and gives us a scope to work. Thus before releasing this kind of data in public government should pay little attention on the data set and provide users with all information.
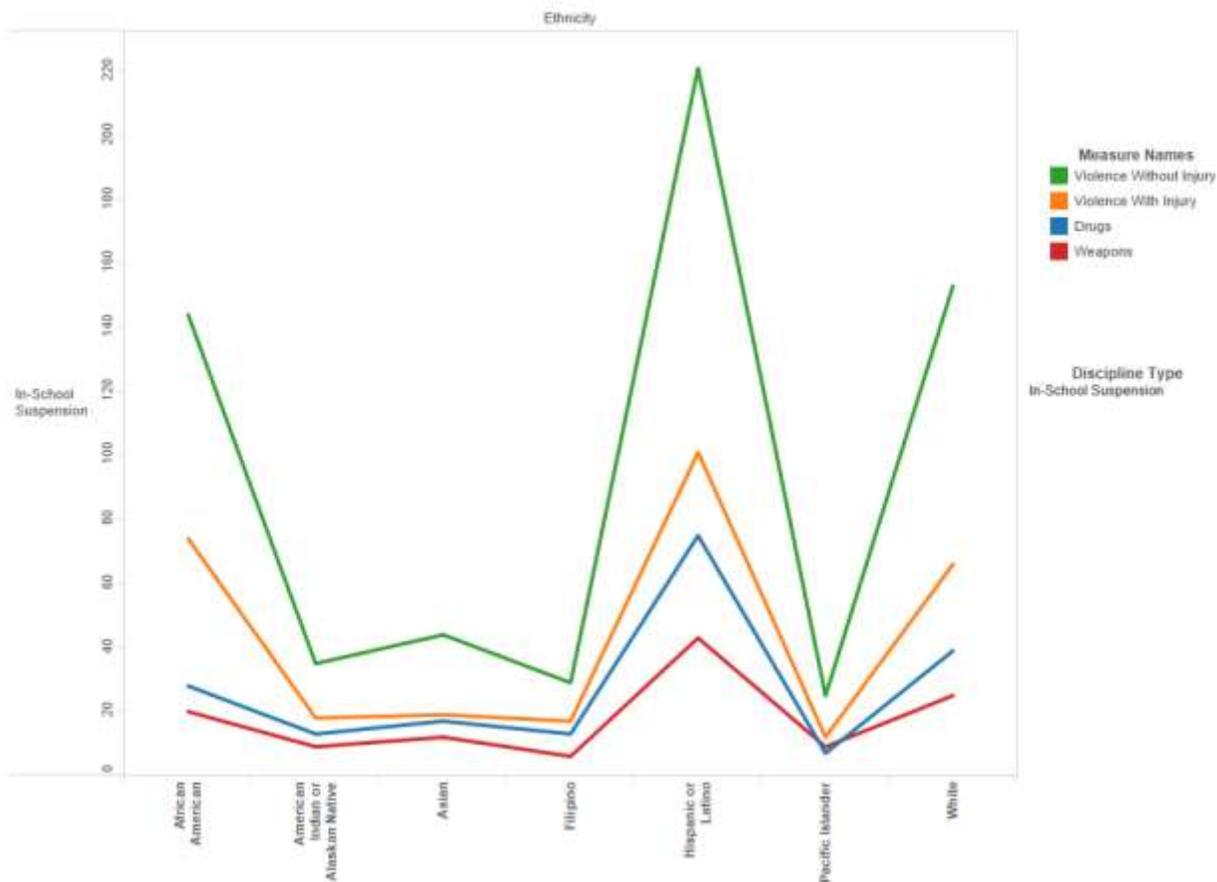
Description: It shows reason for suspension and its type like Expulsion, In school Suspension and out – of –school Suspension. Where reason of suspensions may be like violence without injury, violence with injury, drugs and weapons.
Visualizations Pattern: Multi Line chart, Spot light, dynamic query on Discipline Type and measure names.

Conclusion: We see that education givers take it seriously while making statements about expulsion and suspension. As we see expulsion rate is really low as it's serious and can cost bad career for the student, ruining life.
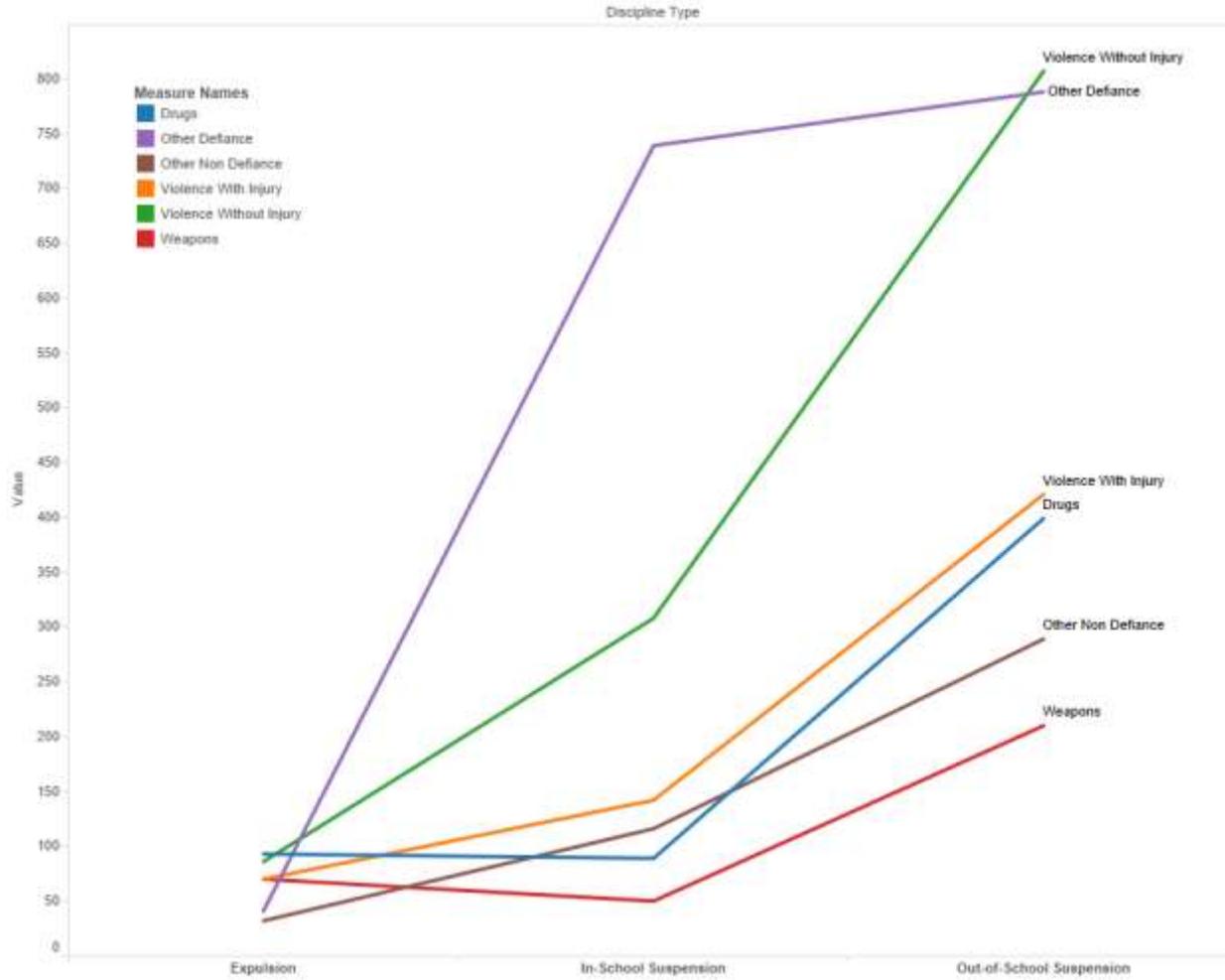
Discipline Type

Measure Names
- Drugs
- Other Defiance
- Other Non Defiance
- Violence With Injury
- Violence Without Injury
- Weapons

Description: It shows reason for suspension and its type like Expulsion, In school Suspension and out – of –school Suspension vs the time line for every graph. Where reason of suspensions may be like violence without injury, violence with injury, drugs and weapons.

Visualizations Pattern: Multi bar chart, Spot light, dynamic query on Discipline Type and measure names

Conclusion: It's just to compare the crime over year and thus we see that effective laws has resulted in decline crime rate.
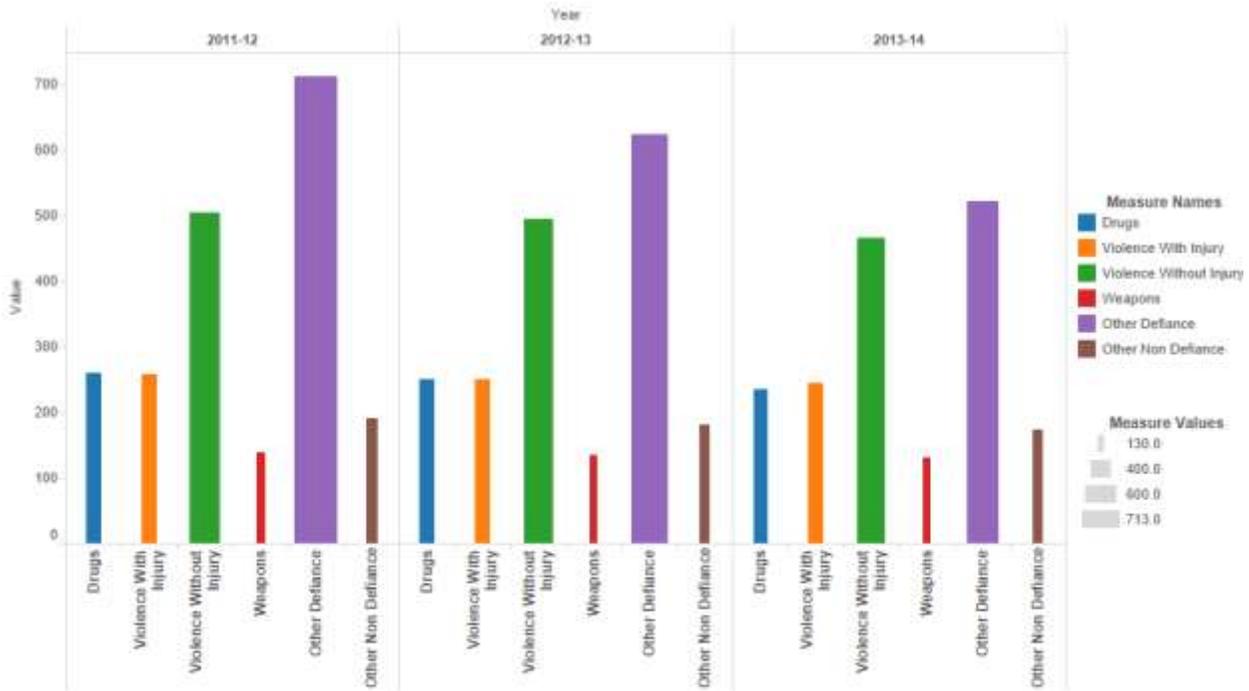
Description: One map shows total number of enrollment and dropout by regions of California. One map shows relative percentage of dropouts by regions of California for comparison. The darker colored regions indicate higher rate of dropout than light colored regions.

Visualizations: Map , Data Tips, Data Spotlight, Dynamic Queries, quick filter, spot light, tooltip.

Conclusion: We see that even the enrollment rate is higher in LA it does not have highest drop out percentage.
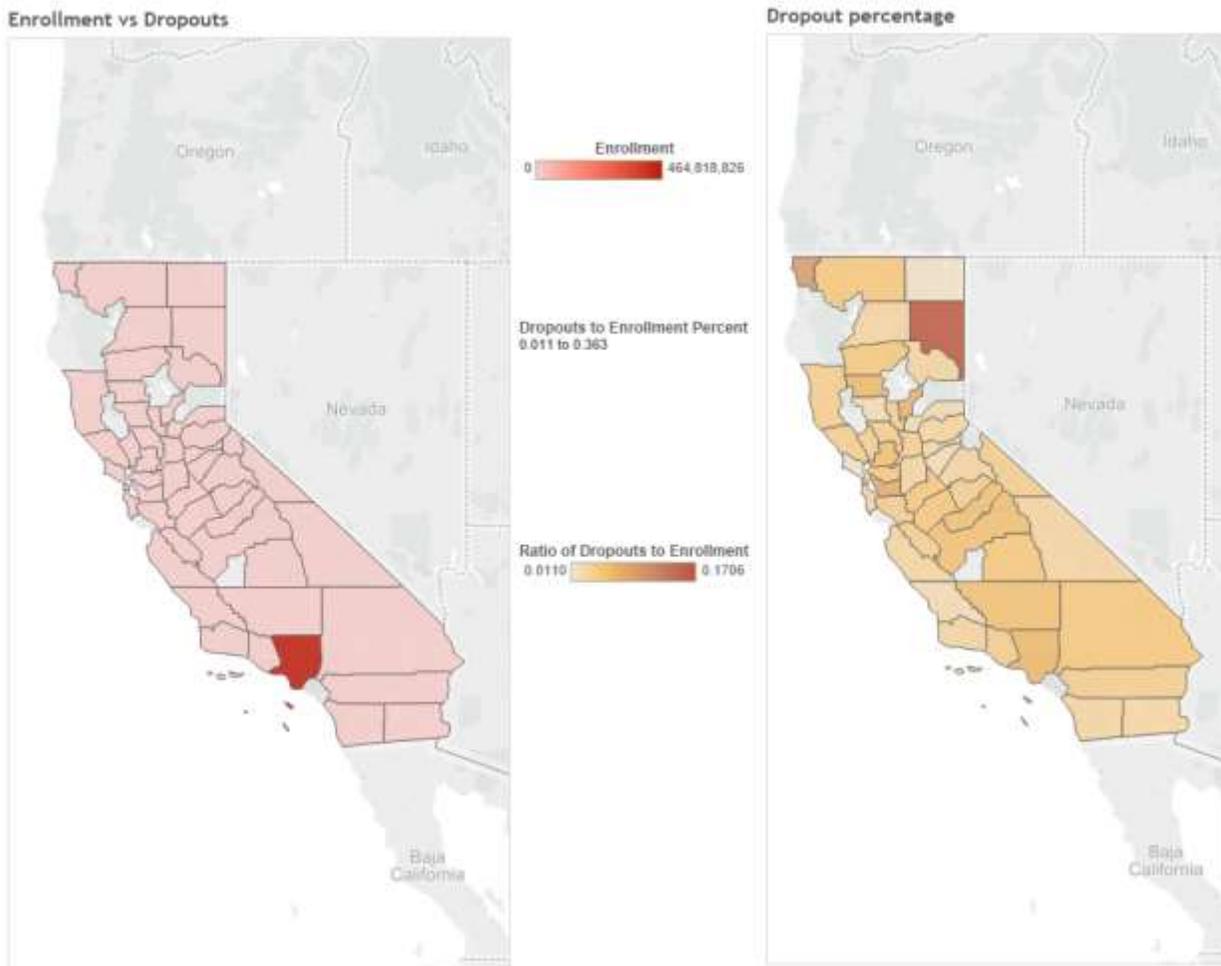
Description: Map shows diversity of languages across multiple regions in California, and how it could affect educational policies controlled by local board of education and state legislation.

Visualizations: Map , Data Tips, Data Spotlight, Dynamic Queries, quick filter, spot light, tooltip.
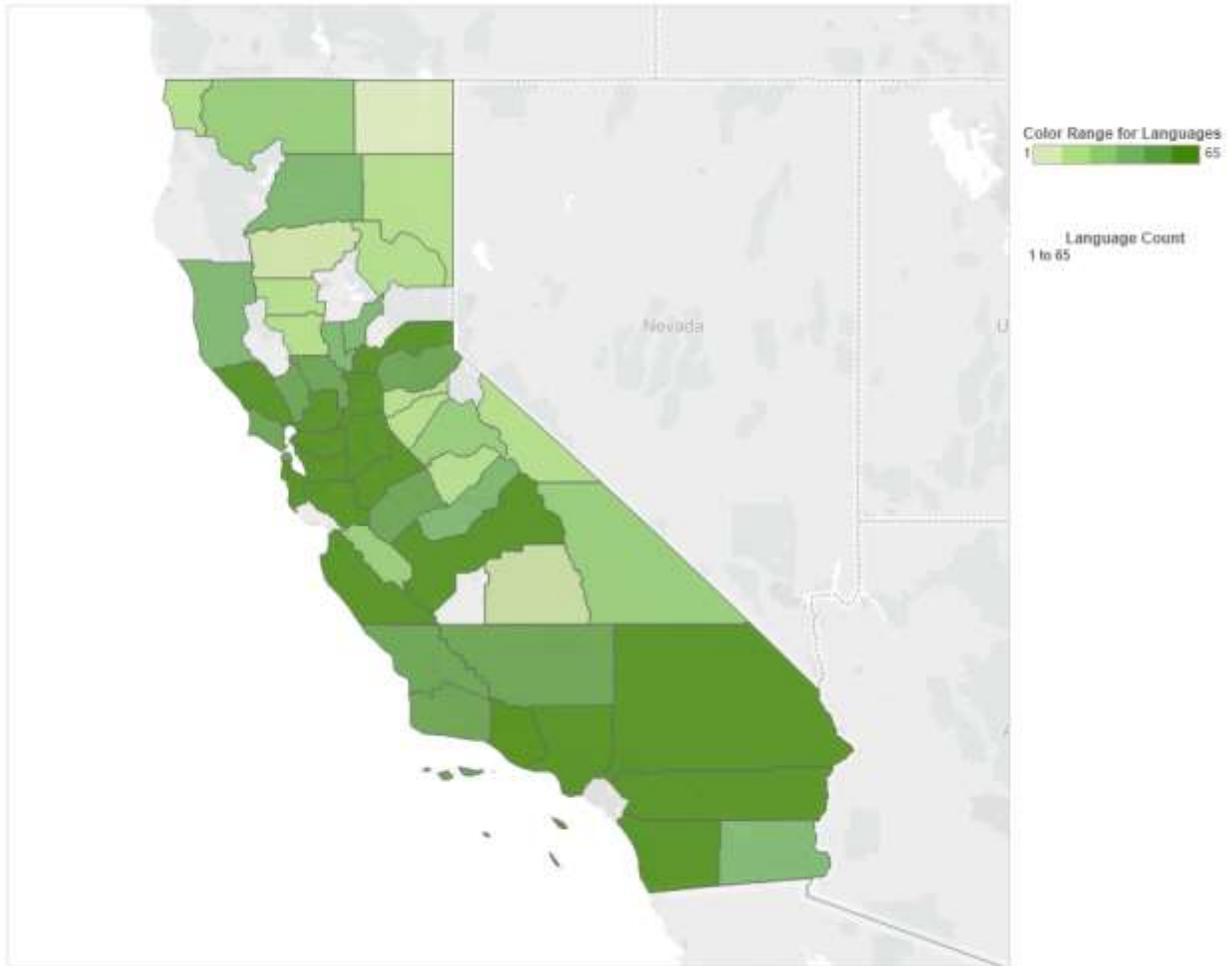
Description: Bar chart show popularity of non-English speaking languages across schools in California, this has high effect on educational legislation

Visualizations: Bar Chart , Data Tips, Data Spotlight, tooltip.
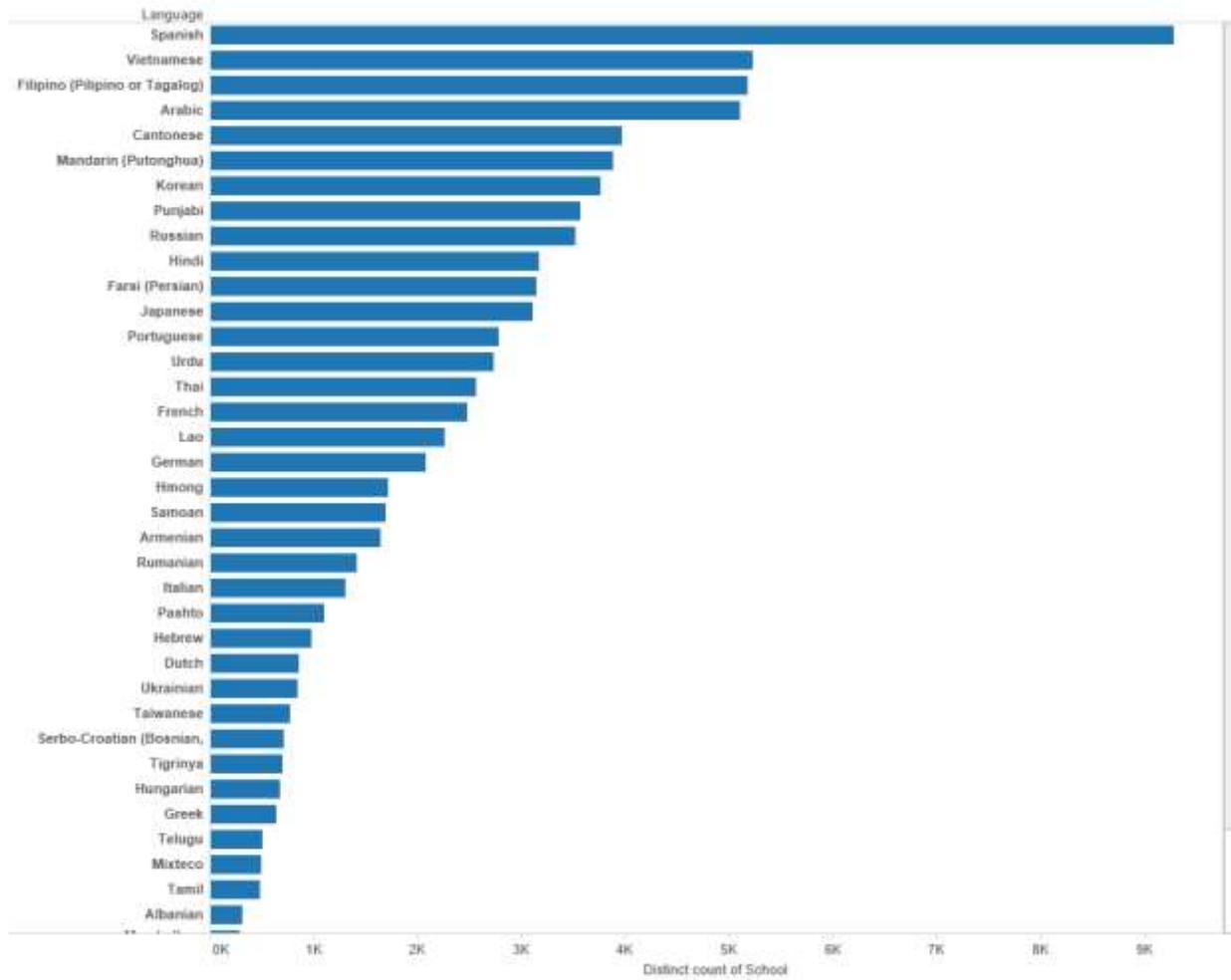
Description: Poverty chart shows relationship between education and family economic hardship throughout the years.

Visualizations Pattern: Line Chart, Tooltip, quick filter, High Charts
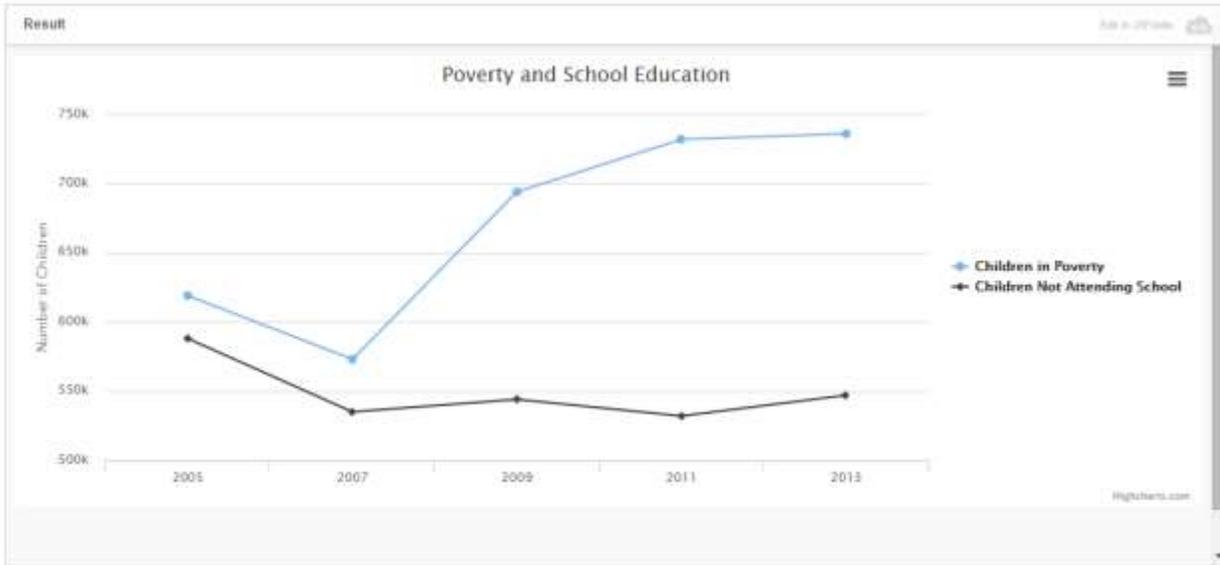
**Description:** Transportation chart shows preferred method of commuting to and from school by students based on their living distance from the school.

**Visualizations Pattern:** Dot plot, multivariate displays: Glyphs
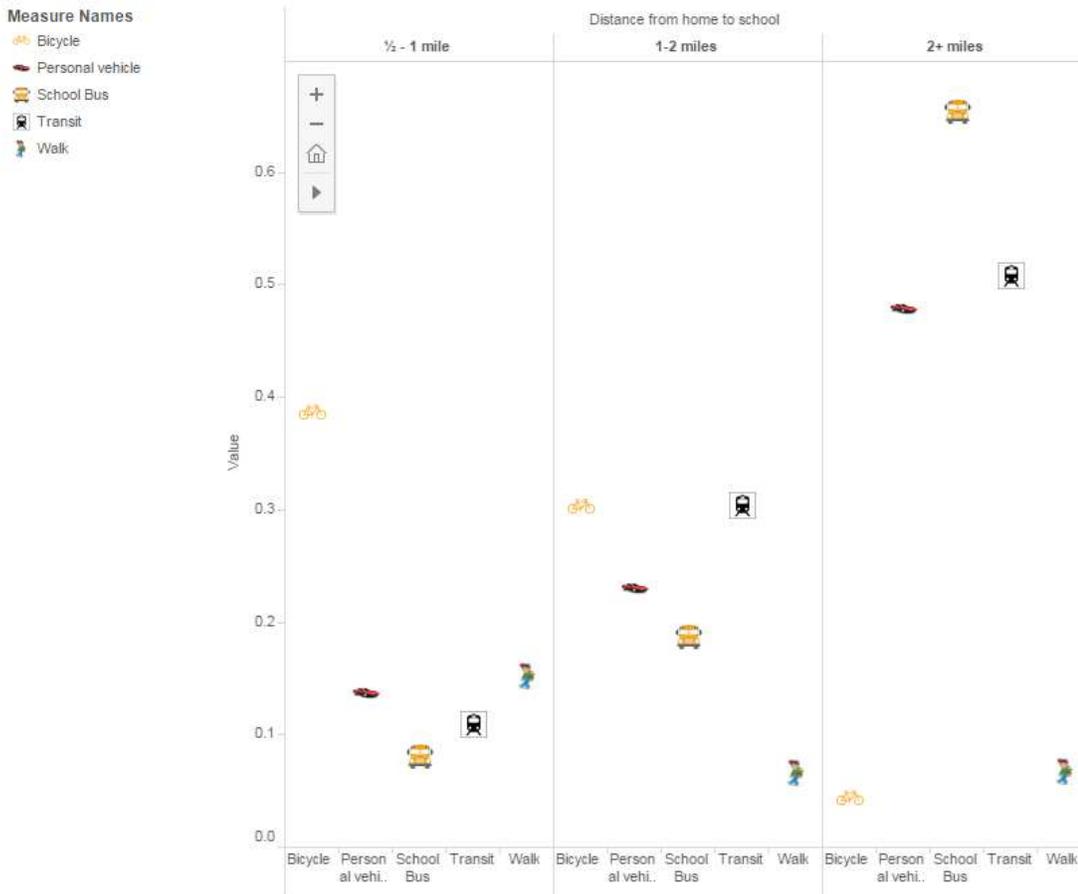
Description: Sport chart shows popular sport in high school and student participation
Visualizations Pattern: Combination chart of bar chart, line graph and pie chart.
We have combination graph and all the views is shown here. The last is the final view with all the visualization.
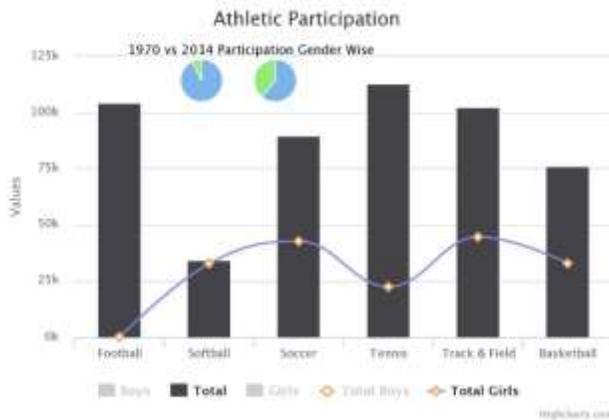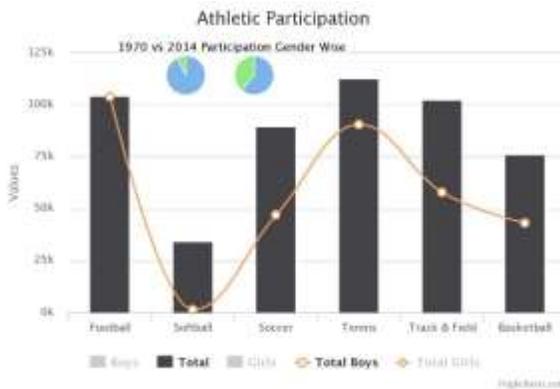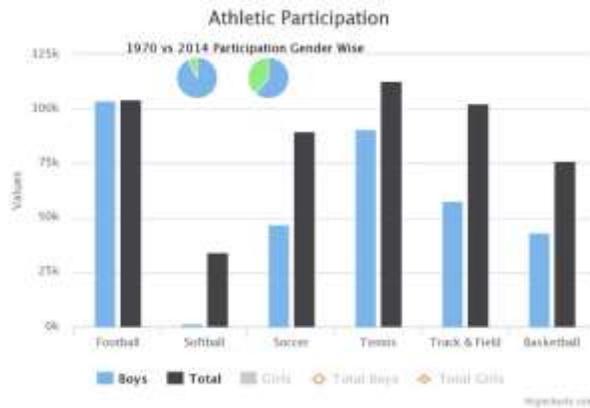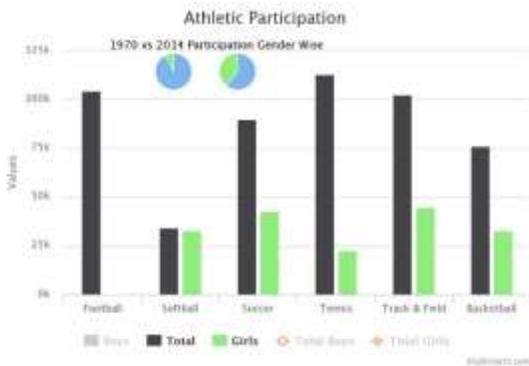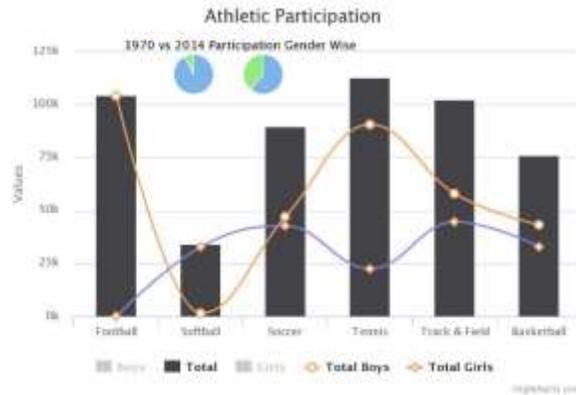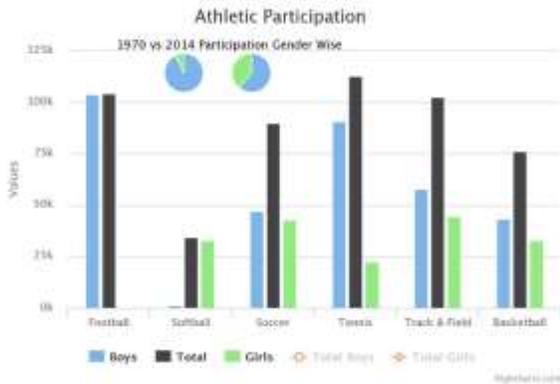
Athletic Participation
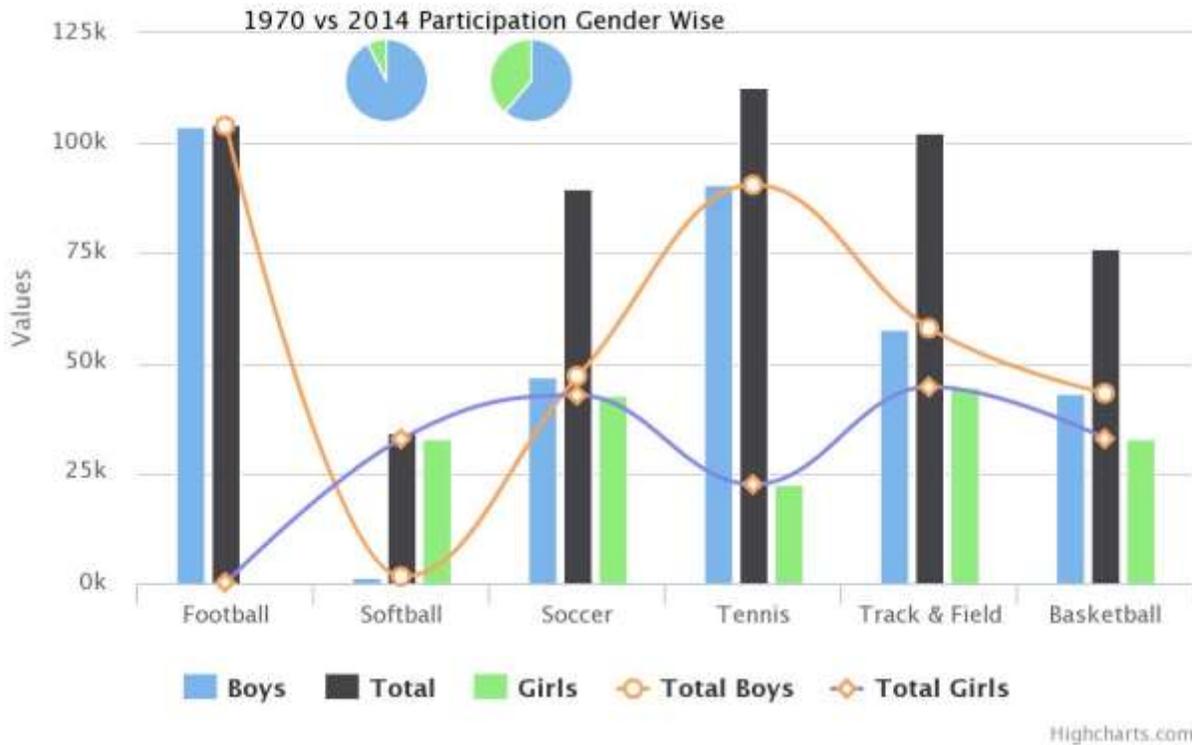
How to Run The Application

As we made our application to be a blog so that it can be of some help to the user. It can be found on the following web Address below.

http://milantechnicalblog.blogspot.com/p/education-data-visulization.html