

1. Data Set

The data set I am using is from a research study done from 2008-2009 called *A Socioeconomic and Recreational Profile of Surfers in the United States*, by Surf-First and The Surfrider Foundation.

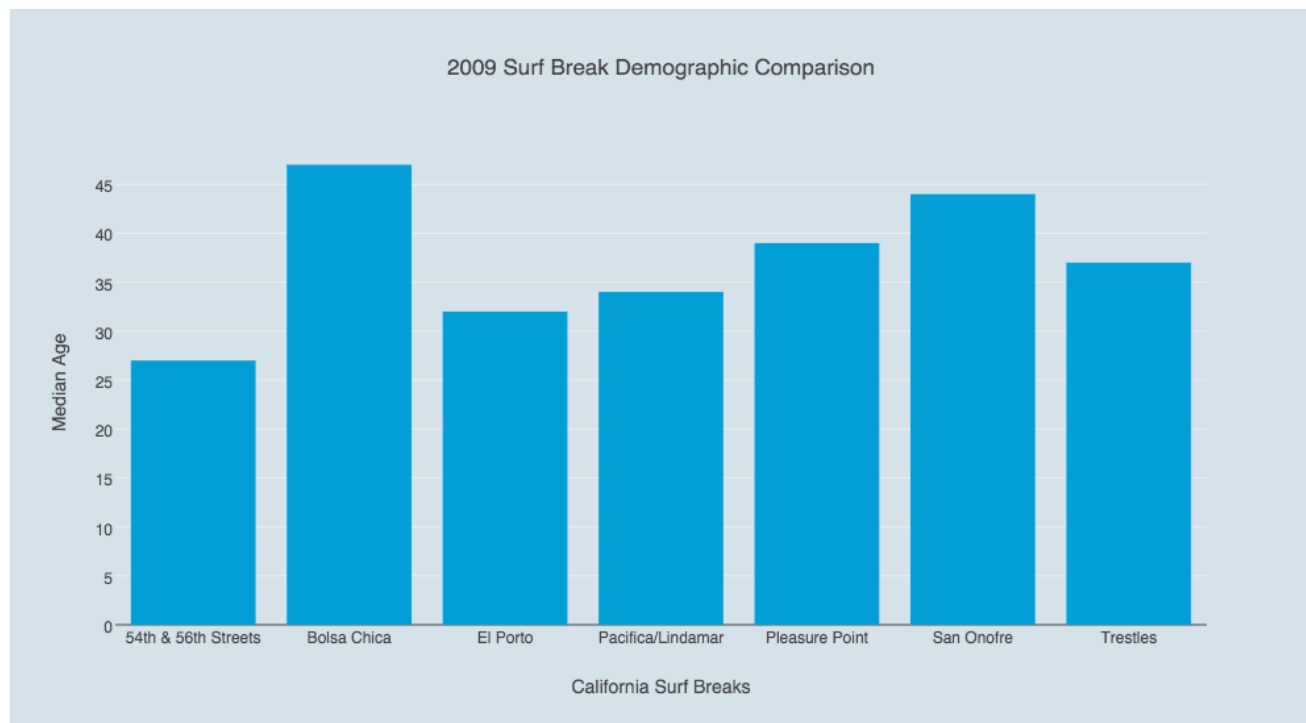
[public.surfrider.org/files/surfrider\\_report\\_v13.pdf](http://public.surfrider.org/files/surfrider_report_v13.pdf)

2. Initial Question

Is there a correlation between age and choice of surf break in major California surf spots?

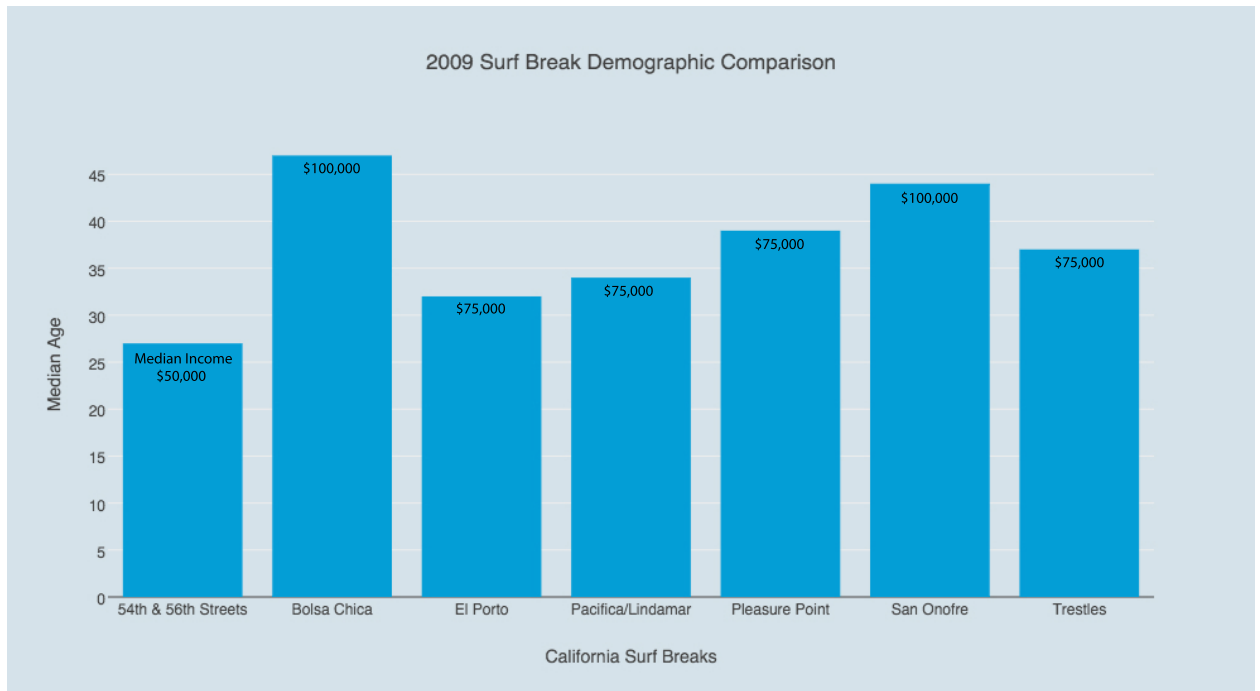
3. Software/Altering

First I dropped the data into an Excel file to pair down the 2-3 variables I wanted to look at specifically (they had 6 or 7 demographic variables listed). Then I dropped the csv file into Plot.ly and got my first bar chart visualization of an x&y axis of surf break and age.

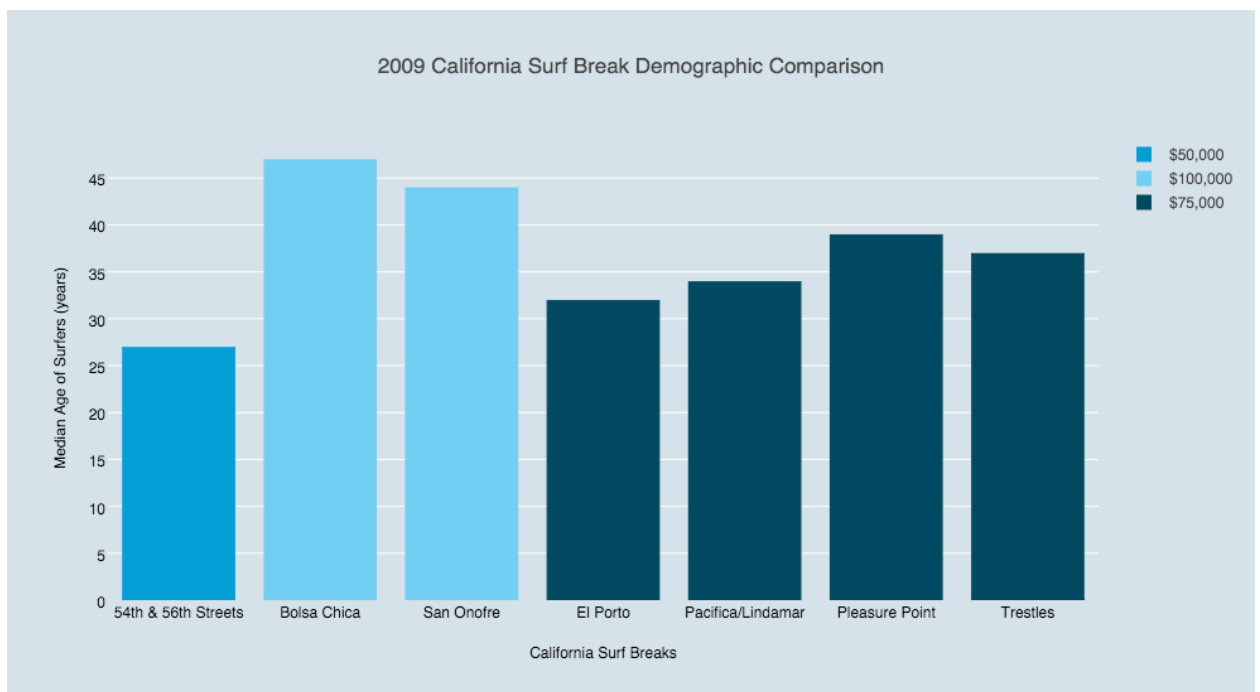


After looking at this bar chart and data, I wanted to know *why* these were the median ages. Being from California helps in assuming that possibly because of where the Universities are can play a factor in the 54<sup>th</sup> & 56<sup>th</sup> streets data point but that was my only initial assumption.

I decided to drop this first visualization into Photoshop and to look at a third variable of median income. I added this variable by placing text at the top of each bar specifying the median income of each group. This also altered my question to: What is the correlation between income and choice of surf break?



I decided this looked unclear and a bit confusing with having the text at the top of each bar because it read as if that was the value of the x-axis. So, I put the Photoshop file back into Plot.ly to represent the 3<sup>rd</sup> variable in another way, ie; color.



This is my final visualization.