

Audrey, Gemi & Vidi
 Mr. Holt
 Statistics M/W 6pm
 17 April 2017

Is Generosity Explained By Income?

Every year, Center For A Non Violent Community (CNVC), an agency in Tuolumne County which provides services for survivors of domestic violence and sexual assault, sends out a Holiday Letter Campaign requesting donations. This letter is sent to every address in Tuolumne County via the Post Office pre-sorted mail service. *Our question is: Does the amount donated per zip code correlate with the Median income of each zip code – indicating that generosity can be explained by income?*

DATA COLLECTION:

We first gathered the Donation income for the last four years (2013-2016)—this information was provided by CNVC—and totaled up the income per zip code from Tuolumne County only. Next, we went on the Post Office Website to find the **number of addresses per zip code** in Tuolumne County. We divided the Total amount of income per zip code by the number of addresses per zip code to come up with our “**Average Donation per household/per Zip Code.**” **SEE BELOW FOR CHART 1:**

City	Zip	# of Addresses	Donations By Zip Code	Average donation per household
Big Oak Flat	95305	249	\$ 325.00	\$ 1.31
Chinese Camp	95309	115	\$ 15.00	\$ 0.13
Columbia	95310	1546	\$ 7,025.00	\$ 4.54
Groveland	95321	2381	\$ 2,370.00	\$ 1.00
Jamestown	95327	3382	\$ 6,095.00	\$ 1.80
Long Barn	95335	290	\$ 320.00	\$ 1.10
MiWuk	95346	787	\$ 2,565.00	\$ 3.26
Pinecrest	95364	176	\$ 350.00	\$ 1.99
Sonora	95370	12000	\$ 36,812.00	\$ 3.07
Soulsbyville	95372	1365	\$ 5,655.00	\$ 4.14
Standard	95373	1517	\$ 2,470.00	\$ 1.63
Strawberry	95375	105	\$ 850.00	\$ 8.10
Tuolumne	95379	1984	\$ 6,455.00	\$ 3.25
Twain Harte	95383	2746	\$ 12,583.00	\$ 4.58

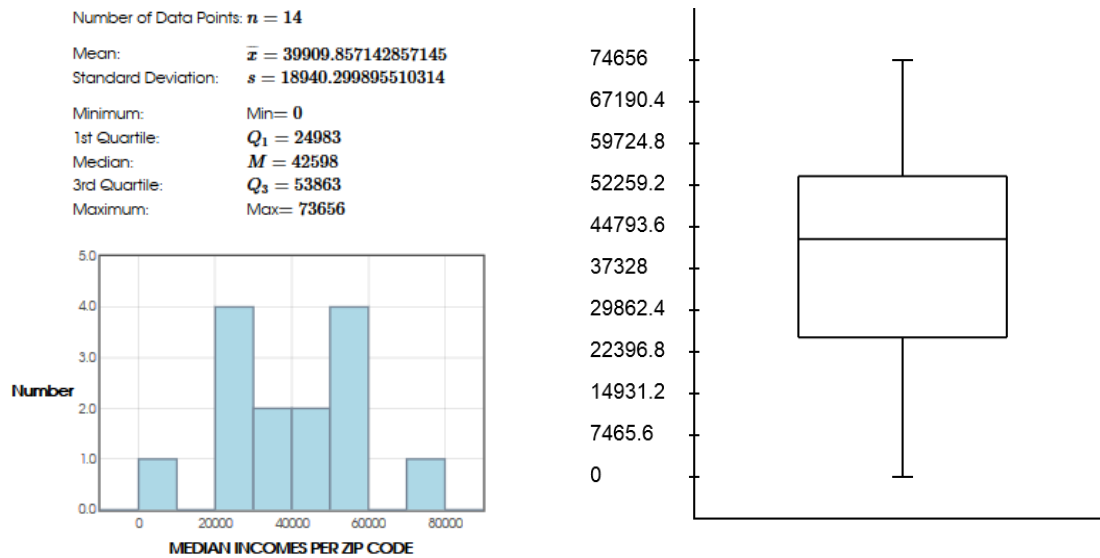
The next set of data we collected was the Median Income per Zip code. This information was a little bit harder to come by as some websites did not provide median income for every zip code in Tuolumne County. We were unable to locate a median income for Standard, 95373, with any website. To get the majority of our data we used “Incomebyzipcode.com” which uses U.S. census to formulate their data. The other websites used were as noted on the chart below: **SEE BELOW FOR CHART 2:**

City	Zip	Median Incomes	Website Used
Big Oak Flat	95305	\$ 25,938.00	www.neighborhoodlink.com
Chinese Camp	95309	\$ 55,140.00	www.incomebyzipcode.com
Columbia	95310	\$ 39,487.00	www.incomebyzipcode.com
Groveland	95321	\$ 53,863.00	www.neighborhoodlink.com
Jamestown	95327	\$ 35,268.00	www.incomebyzipcode.com
Long Barn	95335	\$ 22,147.00	www.neighborhoodlink.com
MiWuk	95346	\$ 54,758.00	www.incomebyzipcode.com
Pinecrest	95364	\$ 24,983.00	www.cenury21.com
Sonora	95370	\$ 49,288.00	www.incomebyzipcode.com
Soulsbyville	95372	\$ 73,656.00	www.incomebyzipcode.com
Standard	95373	\$ -	Cannot find
Strawberry	95375	\$ 24,983.00	www.century21.com
Tuolumne	95379	\$ 53,518.00	www.incomebyzipcode.com
Twain Harte	95383	\$ 45,709.00	www.incomebyzipcode.com

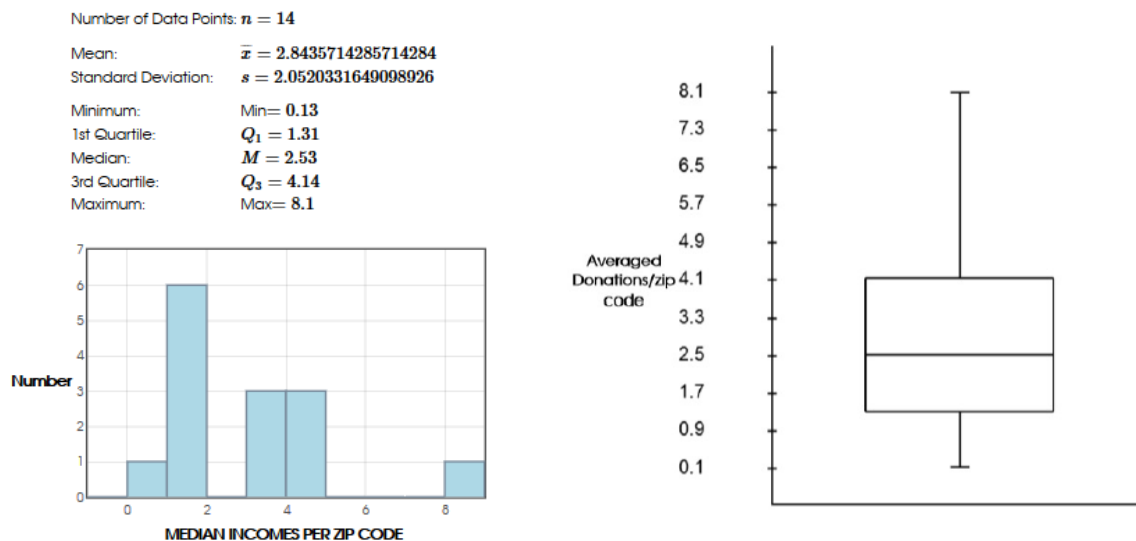
ANALYSIS

The first charts we made (using Holt.Blue Statistical Software) were boxplots of both data sets. We found this helpful in identifying outliers and seeing if we had Normal or Skewed Distributions. The Box Plot for the Median Income showed a Normal Distribution with only one obvious outlier – which was for Standard with a \$0 median income as we were unable to locate it. (See below for GRAPH 1) The Box plot for the Average Donation Data Set showed a skewed distribution to the right, with two obvious outliers: Chinese Camp with an average donation of \$0.13 and Strawberry with an average donation of \$8.10. Both of these outliers can be explained by the small number of addresses for those particular zip codes. (See below for GRAPH 2)

GRAPH 1: Box Plot for Median Incomes per Zip Code in Tuolumne County



GRAPH 2: Box Plot for Average Donation per address per Zip Code in Tuolumne County:



Next, we assigned the data the graph variables “Y” and “X”. We assigned the data set: “Average donation per address per Zip code in Tuolumne County” to the “Y” variable, as this data was the “Response” to be “explained” by the “X” variable the “Median Income per Zip code.” Then, using Statistical Software by Holt.Blue - Correlation and Regression Calculator - we made our first graph which included all outliers. We saw a very weak correlation (almost none at all)

between donation amounts and median income. The correlation was $r=0.081$ and r-squared was 0.006. **SEE GRAPHS 3 & 4 BELOW:**

GRAPH #3:

Correlation and Regression Calculator

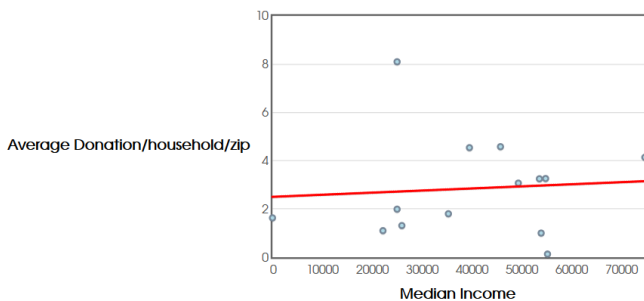
Variable Names (optional): Median Income Donation/household/zip

Explanatory (x) Response (y)

25938	1.31
55140	0.13
39487	4.54
53863	1.00
35268	1.8
22147	1.10
54758	3.26
24983	1.99
49288	3.07
74656	4.14
0	1.63
24983	8.10
53518	3.25
45708	4.58

Data goes here (enter numbers in columns):

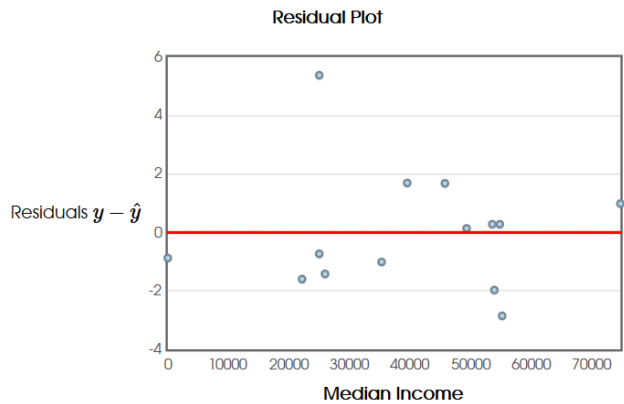
Include Regression Line: Include Regression Inference:



[Download Scatter Plot JPEG](#)

Regression Line: $AVERAGE\ DONATION/HOUSEHOLD/ZIP = 0 \cdot MEDIAN\ INCOME + 2.502$
 Correlation: $r = 0.081$
 R-squared: $r^2 = 0.006$

GRAPH #4: Residual Plot



While we noticed a positive slope showing as the median income raised – so did the average amount donated per address, the correlation was very weak and not strong enough to

come to the conclusion that the Response and Explanatory Variables are related. Another problem is that the way Holt.Blue calculates the slope: there is too much round-off, making the slope of the line 0 when it clearly is positive. Using our graphing calculator, we calculated the slope of the line to be: .0000087 (or $y = .0000087(x) + 2.50$). Even still, we decided to try again without the obvious outliers to see if we would come to a different conclusion.

See below for **CHART 3, GRAPHS 5, 6, 7, and 8:**

CHART 3 : Located Outliers and Removed:

City	Zip	# of Addresses	Donations /Zip	Average donation per household	Zip	Median Incomes
Big Oak Flat	95305	249	\$ 325.00	\$ 1.31	95305	\$ 25,938.00
Chinese Camp	95309	115	\$ 15.00	\$ 0.13	95309	\$ 55,140.00
Columbia	95310	1546	\$ 7,025.00	\$ 4.54	95310	\$ 39,487.00
Groveland	95321	2381	\$ 2,370.00	\$ 1.00	95321	\$ 53,863.00
Jamestown	95327	3382	\$ 6,095.00	\$ 1.80	95327	\$ 35,268.00
Long Barn	95335	290	\$ 320.00	\$ 1.10	95335	\$ 22,147.00
MiWuk	95346	787	\$ 2,565.00	\$ 3.26	95346	\$ 54,758.00
Pinecrest	95364	176	\$ 350.00	\$ 1.99	95364	\$ 24,983.00
Sonora	95370	12000	\$ 36,812.00	\$ 3.07	95370	\$ 49,288.00
Soulsbyville	95372	1365	\$ 5,655.00	\$ 4.14	95372	\$ 73,656.00
Standard	95373	1517	\$ 2,470.00	\$ 1.63	95373	\$ -
Strawberry	95375	105	\$ 850.00	\$ 8.10	95375	\$ 24,983.00
Tuolumne	95379	1984	\$ 6,455.00	\$ 3.25	95379	\$ 53,518.00
Twain Harte	95383	2746	\$ 12,583.00	\$ 4.58	95383	\$ 45,709.00

The highlighted Zip Codes were removed as outliers. The remaining data was entered into the Statistical Software by Holt.Blue “Correlation and Regression Calculator”:

See below for **GRAPH 5:**

Correlation and Regression Calculator

Variable Names (optional):

Median Income Donation/household/zip

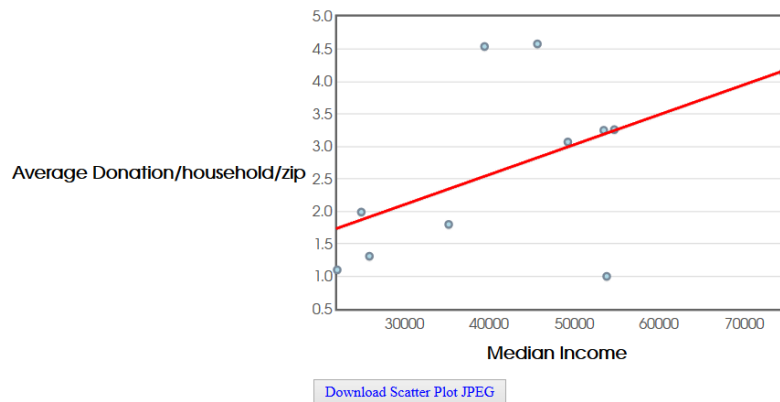
Explanatory (x) Response (y)

Data goes here (enter numbers in columns):

25938	1.31
39487	4.54
53863	1.00
35268	1.8
22147	1.10
54758	3.26
24983	1.99
49288	3.07
74656	4.14
53518	3.25
45708	4.58

Include Regression Line:

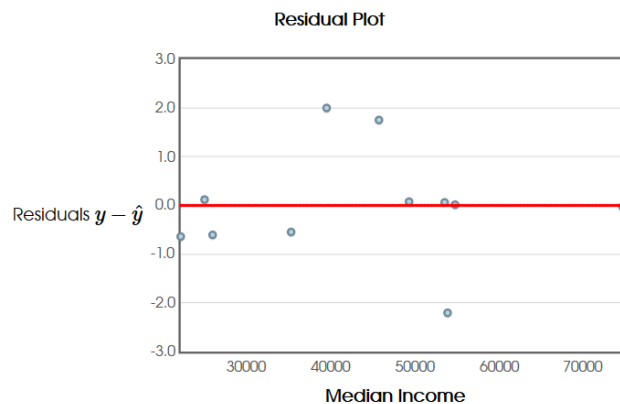
Include Regression Inference:

GRAPH 6:

Regression Line: $\text{AVERAGE DONATION/HOUSEHOLD/ZIP} = 0 \cdot \text{MEDIAN INCOME} + 0.717$

Correlation: $r = 0.542$

R-squared: $r^2 = 0.293$

GRAPH 7:

We noticed a stronger correlation between the Average Donation per Zip Code and the Median Income without the Outliers. The correlation was $r=0.542$ and the r-squared was 0.293 . Again, we had a problem with Holt.Blue calculations for the slope, showing the slope of the line 0 when it clearly is positive. Using our graphing calculator, we calculated the slope of the line to be: $.000046$ (or $y=.000046(x)+0.702$). The slope of the line was positive showing that as the Median Income went up, so did the average donation amount.

To further investigate the correlation between income and generosity we performed one more test using Holt.Blue: The Regression Inference. Our Null Hypothesis is that the slope of the line is equal to zero and the Alternate Hypothesis is that the slope of the line is greater than zero:

GRAPH 8Include Regression Line: Include Regression Inference: Null Hypothesis: $H_0 : \beta = 0$ Alternative Hypothesis: $H_a : \beta >> 0$ Significance level: $\alpha = 0.05$ Regression Inference: $y = \alpha + \beta x$ Degrees of Freedom: $df = n - 2 = 9$ Estimate of Slope: $b = 0$ Standard Error Slope: $SE_b = 0$ Regression Standard Error: $s = 1.197$ t -Statistic: $t = 1.906$ 95% Confidence Interval for β : (0, 0) p -value: $p = 0.044$

Interpretation:

Assuming that the true slope is $\beta = 0$, the probability of seeing a test statistic $t = 1.906$ or bigger is 0.044.That is, assuming that there is no straight-line relationship between Median Income and Average Donation/Household/Zip, 4.4% of all similarly collected samples would have a test statistic of $t = 1.906$ or greater.

Conclusion:

Reject the null hypothesis. ($0.044 = p < \alpha = 0.05$)**CONCLUSION**

There is a weak correlation between the Median Income of zip codes in Tuolumne County and the Average Amount donated by each Zip Code. This is not strong enough evidence to conclude that the Median Income directly effects the amount donated in certain Zip codes or people's generosity.

CITATIONS:

1. Data for Income collected through Holiday Letter Campaign for CNVC:
Center For A Non Violent Community Standard Rd, Sonora CA 95370
2. Number of Addresses per Zip Code in Tuolumne County
Unites States Post Office Website: <https://tools.usps.com>
3. Median Incomes Per Zip Code: (as noted in Chart 2)
www.incomebyzipcode.com
www.neighborhoodlink.com
www.century21.com
4. Software for Charts 1, 2, 3: Microsoft Office Excel
5. Statistical Software for Graphs 1-8: Holt.Blue
6. Graphing Calculator used to find slopes of lines: CASIO fx-9750GII