Data Table and Central Tendency

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Year	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Mean CO_2 Data (PPM)	358.64	360.62	362.36	363.47	366.5	368.14	369.4	371.07	373.17	375.78	377.52	379.76	381.85	383.71	385.57

Mean

Is equal to sum of all data, divided by number of cells. So, add cells B4 through P4 and divide by 15.

=SUM(B4:P4)	5577.6	<-- Total of all data values (358.64 + 360.62 + 362.36 + … + 383.71 + 385.57)
=5578/15	371.87	<-- Total, divided by number of values (15), otherwise known as, the mean

Median

Order all the data from least to greatest, then select the middle value.

358.64	360.62	362.36	363.47	366.5	368.14	369.4	371.07	373.17	375.78	377.52	379.76	381.85	383.71	385.57
Greatest -->					<-- Median ^ -->								<-- Least	

Mode

No values are repeated more than once, so there is no mean.

Central Tendency Explanation

The centra tendency is relevant to the topic because it gives a scientist information regarding the trends in the data. The mean offers an average over the years, what value the measurements tend to hover around. The median offers a measure of what value occupies the center of the data range, and which values are above and below it. This can help to establish what a rough center would be on the data range. The mode allows scientists to see which values occur most often, and they can associate those values with certain times or seasons.

Source: ftp://ftp.cmdl.noaa.gov/ccg/co2/trends/co2_annmean_mlo.txt

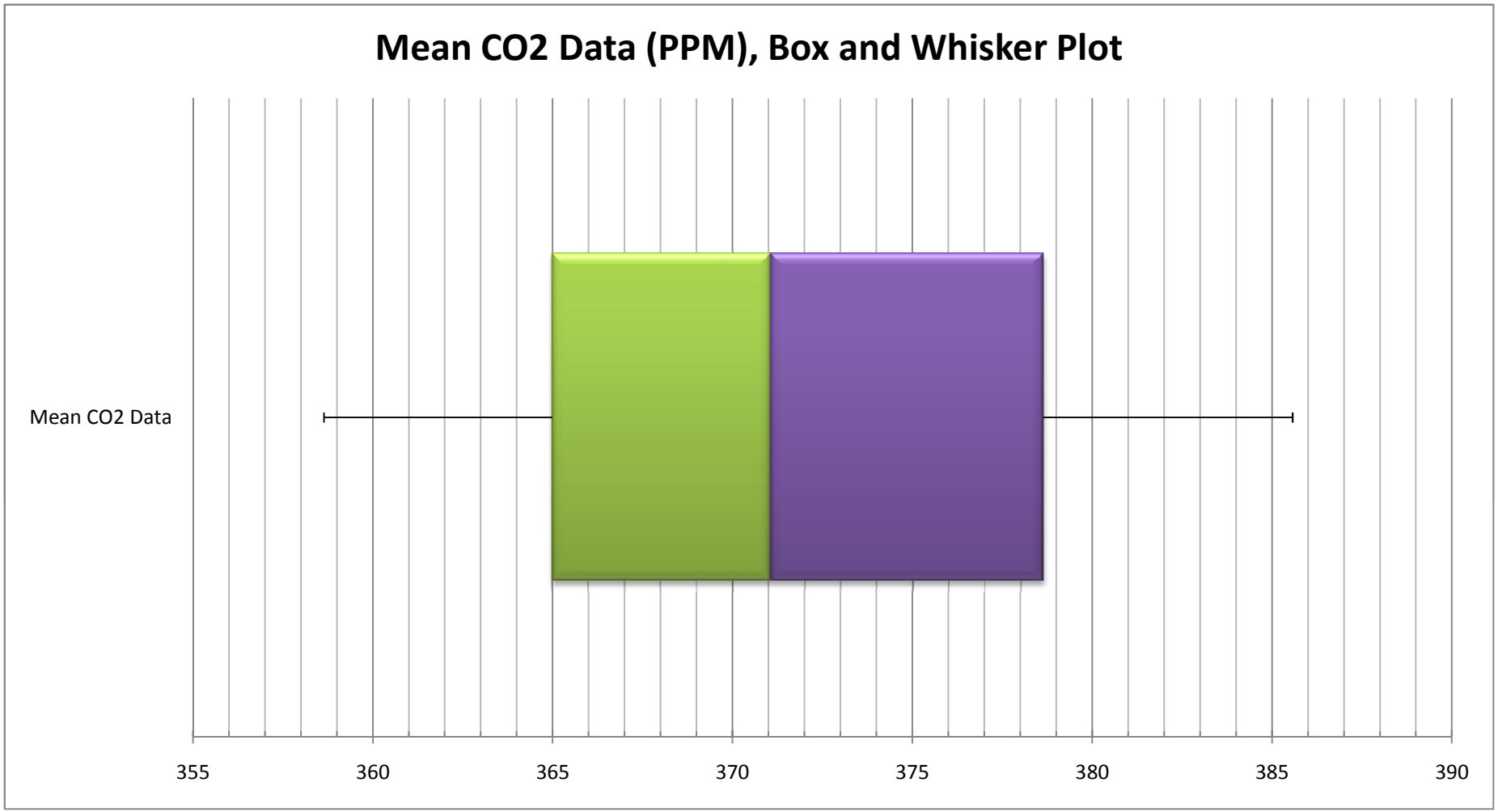
Standard Deviation

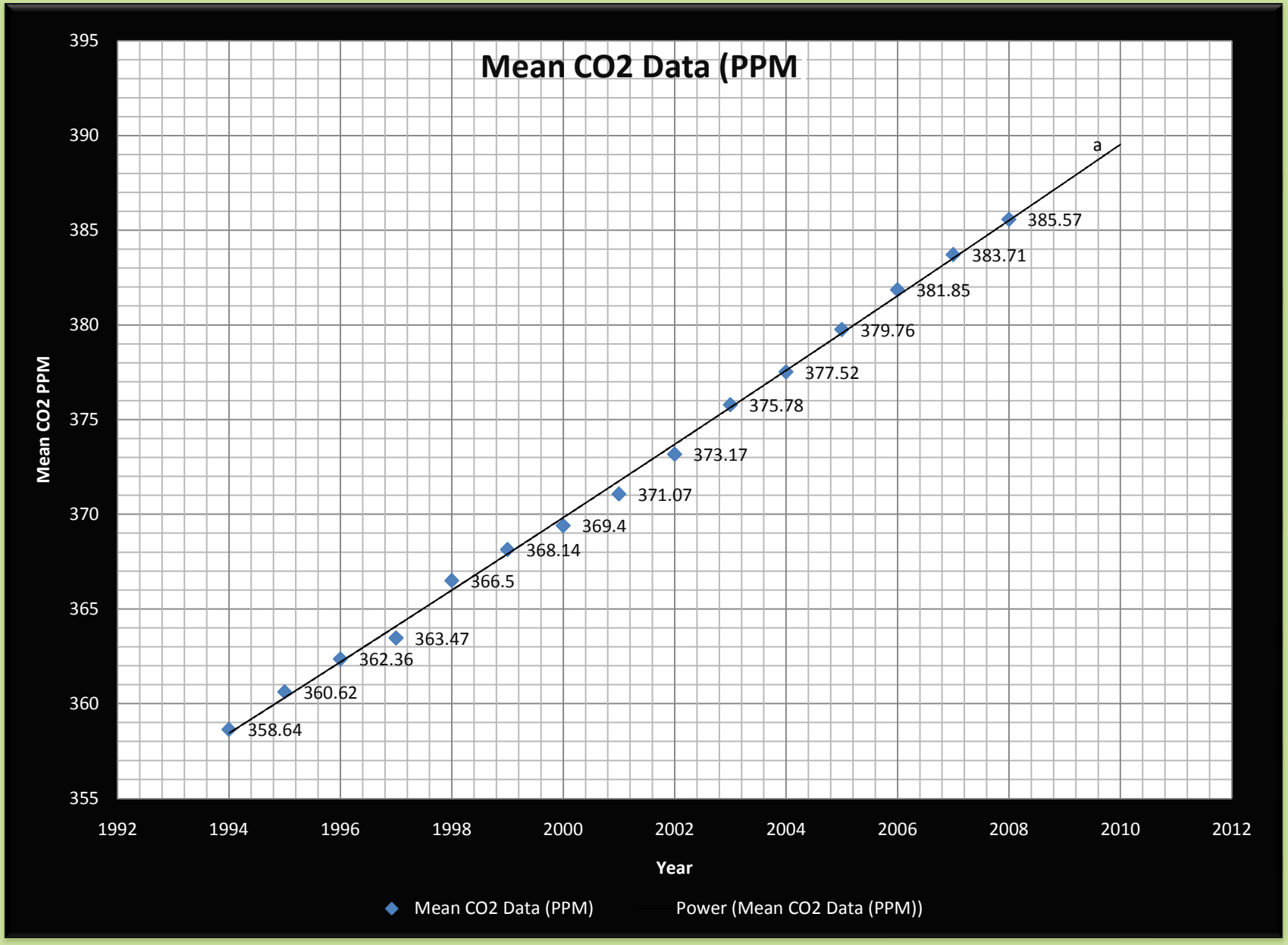
2	358.64	360.62	362.36	363.47	366.5	368.14	369.4	371.07	373.17	375.78	377.52	379.76	381.85	383.71	385.57
3	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667	371.8667
4	13.22667	11.24667	9.506667	8.396667	5.366667	3.726667	2.466667	0.796667	1.303333	3.913333	5.653333	7.893333	9.983333	11.84333	13.70333
5	174.9447	126.4875	90.37671	70.50401	28.80111	13.88804	6.084444	0.634678	1.698678	15.31418	31.96018	62.30471	99.66694	140.2645	187.7813

Step 1	Row 2: List all the values.		
Step 2	Row 3: The value of the mean.		
Step 3	Row 4: The absolute difference from the mean to the coresponding value above it.		
Step 4	Row 5: The square of Row 4 value.		
Step 5	=SUM(A4:O4)	1050.712	Take the sum of all the values from Step 4.
Step 6	=1050.71/15	70.04745	Divide by the quantity of values. This is the variance, or the square of the standard deviation.
Step 7	=SQRT(70.0475)	8.369436	Take the square root for the standard deviation.
Extra	=8.369436*3	25.10831	Three times the standard deviation is the value outside of which, outliers are found.

Outliers

358.64	373.17	369.4	The Median
360.62	375.78	344.2917	Add and subtract amount of three times the standard deviation from the median.
362.36	377.52	394.5083	This gives you the high and low end of the range within which normal values, and not outliers
363.47	379.76		fall.
366.5	381.85		
368.14	383.71	After comparing the high and low range values to my set of values, I can see that no entries fall above or below. Thus,	
369.4	385.57	I have no outliers.	
371.07			


Mean CO2 Data (PPM), Box and Whisker Plot
Mean CO2 Data
355
360
365
370
375
380
385
390


Mean CO2 Data (PPM
395
390
385
380
375
370
365
360
355
Mean CO2 PPM
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Year
a
385.57
383.71
381.85
379.76
377.52
375.78
373.17
371.07
369.4
368.14
366.5
363.47
362.36
360.62
358.64
Mean CO2 Data (PPM)
Power (Mean CO2 Data (PPM))

Harrison Chan

Ms. Green

Algebra 2AB

7 May 2009

Statistics Project: Written Analysis

Box and Whisker Plot

The box and whisker plot essentially separates the data into four data specific sections. At the far left end is the first quartile, containing all the values from value 1 to the value 3.75 on the list. The second and third parts are the thick "box" parts of the plot and contain all the values in the median ranges from the 3.76 mark to the 11.25 value mark. The fourth section is another extremity containing all the values past the 11.26 value. To get the fractional value, a formula is used to calculate what the amount of the fraction is and added to the previous value. Then the values are used to assemble the plot. From the box and whisker plot, it is much easier to see where the widest spreads of values are and where the values are closely grouped. In this data set, a narrow or short section means that CO_2 measurements climbed less during those years. With that knowledge, scientists can study those years to see what reduced our CO_2 output at the time.

Scatter Plot: Correlation Coefficient Relevance

The correlation coefficient, or *r*, is a measure of how well a data set fits the model. In my scatter plot, I used a power regression as the line of best fit. The *r* value came up as .999, meaning that the regression model is an almost perfect fit for the data values. The closer *r* is to 1, the better the values fit the model. The positive value means that there is a positive, or "up", slope on the data set.

Future Predictions

As the data set and the *r* value both indicate, the data has a positive slope. This is bad. As our world becomes more industrialized, we put larger and larger quantities of carbon dioxide into the atmosphere every year. Billions and billions of tons of the gas are released from automobiles, people, factories, and even livestock every year. It has lead to a progressive rise in the overall temperature of the planet by degrees. Scientists have not seen such a rapid rise in all of recorded history. By 2010, CO_2 levels are expect to reach 389.5 PPM. This is an absolutely horrible trend and it needs to be stopped immediately.

We cannot continue to poison our atmosphere and expect to get away with it so easily. Some of the things we can do include reducing our carbon emissions on the road. If you have coworkers, try to carpool whenever possible to reduce the vehicles driving, or even better, use public transit. If you must drive alone, then make sure that your car is well taken care of, with well inflated tires, good oil, and a clean smog test. At home, set your thermostat to 78 during the winter and 72 during the winter. Try to purchase food made closer to home to reduce the "food miles" it has to travel.