

## Extension Statistics Practice #1

### The situation

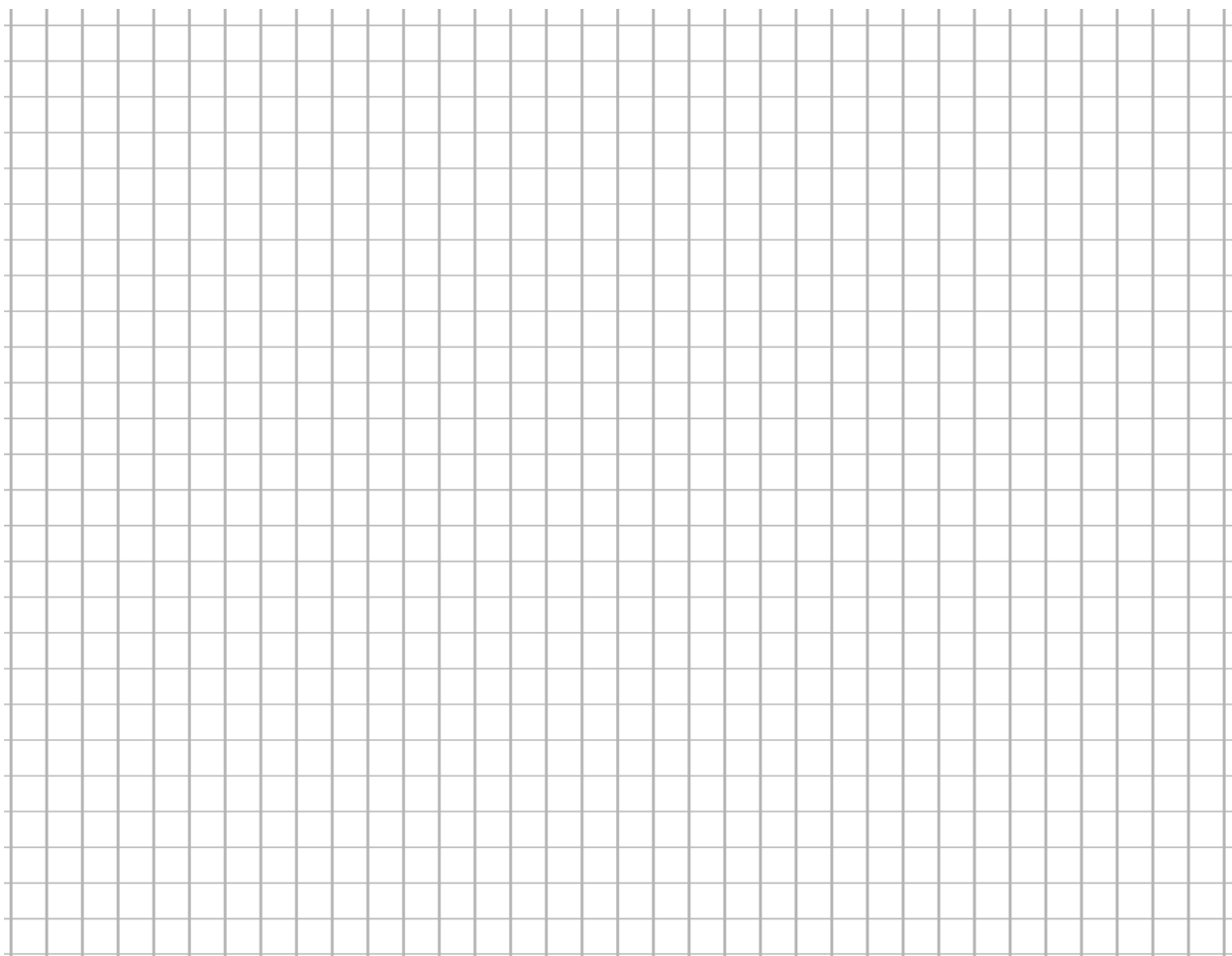
Your local newspaper editor has asked you to investigate the change in house values in his town between 2008 and 2010.

The town has three distinct areas: Waterfront (along the beach), Central (the old town around the main street) and a New Development.

Price data can be found for 2008 and 2010 for a large number of the houses valued or sold in both 2008 and 2010. It is on the next sheet.

### You need to:

- a) Decide the criteria for your report.
- b) Choose appropriate statistics.
- c) Graph the data in an appropriate manner.
- d) Analyse your data and graph
- e) Suggest problems and possible improvements
- f) Write a brief conclusion.



Suburb		2008	2010	dif	Suburb		2008	2010	diff	Suburb		2008	2010	diff
Waterfront	1	537,000	526,000	-11,000	Central	1	586,000	562,000	-24,000	New Dev	1	288,000	223,000	-65,000
Waterfront	2	488,000	483,000	-5,000	Central	2	527,000	503,000	-24,000	New Dev	2	326,000	277,000	-49,000
Waterfront	3	537,000	532,000	-5,000	Central	3	337,000	315,000	-22,000	New Dev	3	344,000	296,000	-48,000
Waterfront	4	559,000	556,000	-3,000	Central	4	421,000	402,000	-19,000	New Dev	4	315,000	268,000	-47,000
Waterfront	5	678,000	677,000	-1,000	Central	5	498,000	489,000	-9,000	New Dev	5	434,000	387,000	-47,000
Waterfront	6	527,000	528,000	1,000	Central	6	420,000	411,000	-9,000	New Dev	6	375,000	333,000	-42,000
Waterfront	7	672,000	678,000	6,000	Central	7	437,000	429,000	-8,000	New Dev	7	538,000	499,000	-39,000
Waterfront	8	544,000	551,000	7,000	Central	8	449,000	441,000	-8,000	New Dev	8	426,000	389,000	-37,000
Waterfront	9	517,000	526,000	9,000	Central	9	443,000	435,000	-8,000	New Dev	9	371,000	336,000	-35,000
Waterfront	10	475,000	486,000	11,000	Central	10	269,000	263,000	-6,000	New Dev	10	311,000	279,000	-32,000
Waterfront	11	439,000	452,000	13,000	Central	11	361,000	356,000	-5,000	New Dev	11	360,000	328,000	-32,000
Waterfront	12	660,000	675,000	15,000	Central	12	515,000	510,000	-5,000	New Dev	12	312,000	285,000	-27,000
Waterfront	13	632,000	649,000	17,000	Central	13	321,000	318,000	-3,000	New Dev	13	382,000	355,000	-27,000
Waterfront	14	641,000	659,000	18,000	Central	14	558,000	557,000	-1,000	New Dev	14	440,000	413,000	-27,000
Waterfront	15	668,000	687,000	19,000	Central	15	463,000	462,000	-1,000	New Dev	15	383,000	357,000	-26,000
Waterfront	16	563,000	583,000	20,000	Central	16	425,000	425,000	0	New Dev	16	244,000	221,000	-23,000
Waterfront	17	467,000	488,000	21,000	Central	17	305,000	314,000	9,000	New Dev	17	424,000	401,000	-23,000
Waterfront	18	627,000	649,000	22,000	Central	18	453,000	464,000	11,000	New Dev	18	246,000	223,000	-23,000
Waterfront	19	618,000	640,000	22,000	Central	19	327,000	339,000	12,000	New Dev	19	428,000	406,000	-22,000
Waterfront	20	790,000	816,000	26,000	Central	20	429,000	442,000	13,000	New Dev	20	365,000	343,000	-22,000
Waterfront	21	728,000	754,000	26,000	Central	21	440,000	454,000	14,000	New Dev	21	401,000	383,000	-18,000
Waterfront	22	564,000	591,000	27,000	Central	22	508,000	522,000	14,000	New Dev	22	257,000	241,000	-16,000
Waterfront	23	601,000	634,000	33,000	Central	23	607,000	625,000	18,000	New Dev	23	416,000	400,000	-16,000
Waterfront	24	541,000	575,000	34,000	Central	24	359,000	379,000	20,000	New Dev	24	284,000	268,000	-16,000
Waterfront	25	489,000	524,000	35,000	Central	25	578,000	599,000	21,000	New Dev	25	342,000	329,000	-13,000
Waterfront	26	622,000	658,000	36,000	Central	26	384,000	408,000	24,000	New Dev	26	322,000	310,000	-12,000
Waterfront	27	489,000	525,000	36,000	Central	27	313,000	339,000	26,000	New Dev	27	413,000	403,000	-10,000
Waterfront	28	761,000	800,000	39,000	Central	28	418,000	451,000	33,000	New Dev	28	368,000	361,000	-7,000
Waterfront	29	695,000	735,000	40,000	Central	29	613,000	647,000	34,000	New Dev	29	468,000	462,000	-6,000
Waterfront	30	694,000	735,000	41,000	Central	30	600,000	636,000	36,000	New Dev	30	405,000	401,000	-4,000
Waterfront	31	636,000	684,000	48,000	Central	31	422,000	461,000	39,000	New Dev	31	292,000	291,000	-1,000
Waterfront	32	696,000	744,000	48,000			<b>13,788,008</b>	<b>13,960,010</b>	<b>172,000</b>	New Dev	32	414,000	414,000	0
Waterfront	33	475,000	525,000	50,000						New Dev	33	400,000	400,000	0
Waterfront	34	507,000	561,000	54,000						New Dev	34	409,000	411,000	2,000
Waterfront	35	443,000	498,000	55,000						New Dev	35	304,000	308,000	4,000
Waterfront	36	587,000	651,000	64,000						New Dev	36	339,000	349,000	10,000
Waterfront	37	672,000	739,000	67,000						New Dev	37	315,000	337,000	22,000
Waterfront	38	426,000	497,000	71,000								<b>13,463,008</b>	<b>12,689,010</b>	<b>-774,000</b>
		<b>22,267,008</b>	<b>23,273,010</b>	<b>1,006,000</b>										



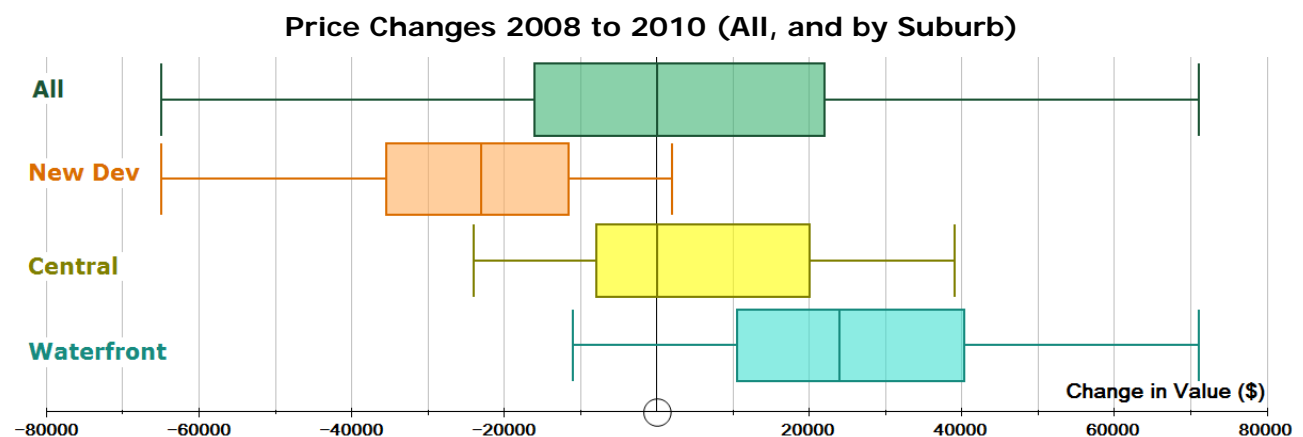
## Model Answers: Extension Statistics Practice #1

(Note: this is particularly complete to show the range of possible answers. Students would not be expected to work at this level.)

I intend to examine the change in house values by the amount the dollar values have differed between 2008 and 2010. I will examine each suburb separately, plus the town as a whole.

I calculated the following statistics:

	Waterfront	Central	New Development
Mean change in value	\$26,474	\$5,548	-\$20,919
Median change in value	\$24,000	\$0	-\$22,000
Biggest decrease	-\$11,000	-\$24,000	-\$65,000
Biggest increase	\$71,000	\$39,000	\$22,000



### Analysis

Both the raw statistics and the graph show that each suburb has a very different pattern of changes. The range for all values suggests little will be gained by looking at the town as a whole.

For the three suburbs the medians are all outside the IQRs of the other boxes, which indicates that the median values are different by a significant amount (that is, outside the range likely to be a statistical variation, so they probably represent true differences in the populations).

The typical house in the Waterfront suburb has increased in price by around \$25,000 (mean = \$26,500, median = \$24,000). There is a large range in differences however (range = \$82,000) in a more or less symmetrical pattern of increases and decreases (seen from each box in the graph being approximately equal) but with a small number of extremely high increases (which act to make the mean about \$2500 more than the median).

Again there is a large range of value differences in the Central suburb (\$63,000). The typical house in Central has not changed much in value, which we can see because by the median change of \$0. However the mean change is +\$5,000 because the houses that increase in value did so by more than the houses that lost value. We can see this in the box and whisker graph, as the upper quartile and upper boxes are larger (more range) than the lower two boxes.

In the New Development (range = \$87,000) all the houses except 6 lost value. The mean and median loss are very similar at \$21,000 and \$22,000. The spread of changes for typical houses in this suburb is tighter than the other two suburbs, which we can see from the much smaller inter-quartile range. One really large loss of \$65,000 really affects the range for New Development, and might be considered an outlier.

## Improvements

We need to know whether the data is representative of all houses in the suburb. If it excludes one group (say auction sales, or mortgagee sales) it might be biased towards high or low values. The best method would be to select houses truly at random from the suburbs and see what the pattern was then.

Rather than dividing by suburb it might be more useful to see how the prices have changed relative to their starting price. Have the cheaper houses lost value and the expensive ones increased? That would involve sorting the prices into groups based on their 2008 value (say bands of \$100,000) and then doing a similar analysis.

It might be interesting to see if there are patterns inside each suburb. Is Central all staying more or less the same, or is part of it increasing and part of it decreasing?

Longer term data might also allow us to conclude whether the patterns are unusual. Has Waterfront been steadily gaining value for a while now?

## Conclusion

It is hard to speak of any change in the typical house value in the town. Values have changed over a very large range (−\$65,000 to +\$71,000), with a fairly even spread of values in between.

Suburb by suburb it is a different story. Waterfront properties have mostly gained value whereas the typical New Development property has lost value. As each of the medians are outside the others' IQR boxes we can have confidence they represent real difference in values between each of the suburbs.

---

## Notes about writing answers:

- 1) Median is the primary statistic and you should start with it. Then comment on ranges and inter-quartile ranges. Means should be used only if you are sure you understand them.
- 2) The comments on the statistics should give both values and their meaning. Try to always put **numbers** to your answers to back up your comments.
- 3) Graphs must be chosen **relevant** to the analysis, not just because they are easy to draw. (Mostly you will find that histograms, bar charts and pie charts are not suitable.)
- 4) Higher level comments should aim to make a **comparison**. That is, while one suburb may be gaining value and another losing, the most interesting comments will compare these two statistical facts in useful ways.
- 5) For Merit try to make references to inter-quartile ranges, extreme values and the symmetry of the distributions. This includes mentioning that nothing unusual is present.
- 6) For Excellence try to make references to the requirements for completeness, generality, representativeness, bias etc of the information used. Is the data – and therefore the statistics calculated – actually valid and useful, or just a bunch of numbers?
- 6) Comments about problems and improvements must be **statistical** in nature. If the comment isn't about the data, graphs or statistics, it probably isn't relevant. This is not Social Studies – comments about why properties lose value, for example, are not relevant.
- 7) Stick to answering the question asked. In this case you are **not** asked which part of town in best value. You will get no marks for analysing something you are not asked to analyse.