# Seeing the Forest with the Stems-and-Leaves

Kirk G. Fleming, FCAS, MAAA

**Abstract:** A picture is worth a thousand words. This paper shows how to use pictures called stem-and-leaf diagrams to display important loss patterns that might otherwise remain hidden in development triangles. These diagrams have the added benefit of appealing to the "big picture" folks in your audience. So that important patterns are always observed in these diagrams, this paper also presents some good practice suggestions that are used to review and evaluate another type of diagram, electrocardiograms (ECG) tracings.

**Keywords**. Reserving, development triangles, stem-and-leaf, ECG, EKG

"You can't see the forest for the trees" is an idiomatic expression that has been around for at least 500 years. It means that someone gets so engrossed in the details that they can't see the big picture. And the reverse expression is also used, albeit less frequently, "You can't see the trees for the forest." Here the expression is highlighting the danger of drawing broad general conclusions about something while missing important information that might be contained in the details.

When dealing with actuarial data, we usually move between two extremes of data detail. At one end we are looking at detailed claims data where we have information on each and every claim evaluated at periodic intervals. At the other extreme we group the individual claims data into development triangles and we look for, measure, and project broad claim trends. But anyone who has been at this for a while knows that the individual claims data is sometimes too much information and the triangles might hide important trends.

In this paper, I would like to offer an alternative way to look at claims data based on the idea of stem-and-leaf displays. Stem-and-leaf displays offer a compromise between the individual claims data and aggregate claim triangle data that presents a chance to observe additional important trends that might otherwise be lost.

So what is a stem-and-leaf display? A stem-and-leaf display is a statistical technique for presenting data where each numerical value is divided into two parts. The leading digit becomes part of the stem and the trailing digit becomes the leaf. The stems are located along the main vertical axis, and the leaves are each observation along the horizontal axis. [1]

As an example, suppose we had the following seven observations between 90 and 100: 96, 95, 93, 96, 97, 98 and 99. The stem for these seven observations would be the 9 and the leaves would be the trailing digits. Setting these up in a stem-and-leaf display we would have the following:

9|3 5 6 6 7 8 9

You just organize the trailing digits from highest to lowest.

The diagram below was created with more data and I am sure you need no further explanation in how it was constructed.

8| 8 9

9| 3 4 6 6 7 8 9

10| 3 3 4 6 7 8

11| 1 2 2 3 3 7 7 8 9

12| 0 0 4 5 5 5 7 7

13| 2 4 5 6 8 9 9

14| 2 3 8

15| 5 5 6

You can see why the diagram is called stem-and-leaf. It looks like a histogram on its side but you have kept all the information about the individual numbers in your collection.

My suggestion is that in addition to using aggregate claims data triangles you should produce stem-and-leaf displays with the emphasis on the word "display." I am going to suggest that you actually deemphasize the numbers for this exercise and just produce pictures that show the additional information that underlies your aggregate claims data.

This organization of your data will allow you to read and tell the story in your data. This presentation will appeal to the "big picture" folks in your organization. You will be able to get your important points across clearly and in a short amount of time.

So let's not beat around the bush and just get right to an example. We can go through the process with some simulated claims data. I simulated five years' worth of claims data between $100,000 and $5,000,000 for this paper. I had a particular story that I wanted to create with this example data. One year has a problem with an increase in the frequency of all sizes of claims. Then things settled down again although we will have a year where some unusually large claims popped up. These changes are the result of changes in claim frequency. The underlying exposure stays the same each year. For this example, I am only going to be using data from the first evaluation column of a development triangle but you could apply this concept to any evaluation column or columns.

The first evaluation column of the simulated data is shown below:

| | |
|---|---|
| 2008 | 19,406,000 |
| 2009 | 21,704,000 |
| 2010 | 41,567,000 |
| 2011 | 36,096,000 |
| 2012 | 23,557,000 |

Rather than break the data into equal buckets as is done with the pure stem-and-leaf display, you should decide how to create the horizontal breaks in your data based on your needs. You might want to create horizontal breakpoints that reflect your definitions of basic and excess limits data. Or you might want to break the data into layers that match provisions in your reinsurance programs. The data should be separated into manageable chunks so that you do not have hundreds of data points on one line and two or three data points on another line. Finally, the splits do not have to be based on the numerical values of the claims but could be based on any type of claim feature.

I am going to use the following stem definitions for my horizontal break points because they fit my data the best and will allow me to highlight some points:

100,000 up to 200,000

200,000 up to 300,000

300,000 up to 500,000

500,000 up to 1,000,000

1,000,000 up to 2,000,000

2,000,000 up to 3,000,000

3,000,000 up to 4,000,000

4,000,000 up to 5,000,000

Enter your data into a spreadsheet using the stem breaks that you selected. If your original selections for breakpoints do not explain your point, you can always go back to the drawing board. I copied a sample of my spreadsheet in Figure 1 on the next page.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | |
| 2 | | 2008 | 100 | 101 | 108 | 110 | 113 | 114 |
| 3 | | | 200 | 201 | 209 | 215 | 219 | 225 |
| 4 | | | 300 | 304 | 307 | 309 | 309 | 332 |
| 5 | | | 500 | 516 | 605 | 618 | 665 | 671 |
| 6 | | | 1,000 | 1,008 | 1,046 | 1,332 | 1,731 | |
| 7 | | | 2,000 | | | | | |
| 8 | | | 3,000 | | | | | |
| 9 | | | 4,000 | | | | | |
| 10 | | | | | | | | |
| 11 | | 2009 | 100 | 100 | 103 | 104 | 112 | 115 |
| 12 | | | 200 | 205 | 220 | 234 | 235 | 243 |
| 13 | | | 300 | 314 | 320 | 343 | 351 | 366 |
| 14 | | | 500 | 514 | 535 | 537 | 555 | 635 |
| 15 | | | 1,000 | 1,054 | 1,180 | 1,189 | 1,561 | 1,739 |
| 16 | | | 2,000 | | | | | |
| 17 | | | 3,000 | | | | | |
| 18 | | | 4,000 | | | | | |

Figure 1

Once you have entered all your data, shrink the page so that you have a much smaller view of the page. In Figure 2, I shrunk my page down to 50%. We cannot read the numbers in the diagram and that is on purpose. We do not care about the numbers at this point. We are looking at the picture.

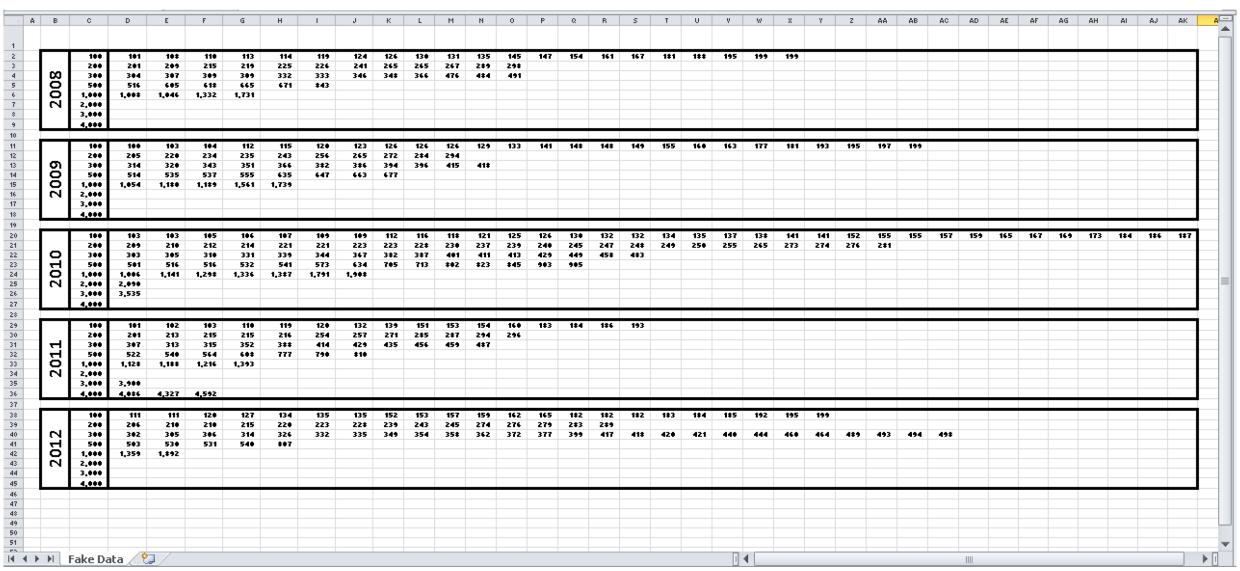

Figure 2

Before I take you through the story in the diagram in Figure 2, I will offer four good practice suggestions adopted from people who are trained to read another type of diagram, an electrocardiogram (ECG). ECG's (or EKG's if you are old school) are diagrams of the electrical activity of the heart captured by putting electrical wires around a person's heart. The heart sends electrical impulses down special internal pathways so that it contracts in a highly coordinate fashion. The typical ECG that is done in the hospital or your doctor's office shows 12 different views of the electrical activity. By looking at variations from the norm in those twelve different pictures, a doctor can diagnose electrical or physical changes in the heart that are causing variations in the pictures.

- The first suggestion is to know what the normal year's picture looks like. Once you know what is normal, then you can spot what is not normal. You might not know what is causing a year to be different but you will know that something deserves more study.
- The second suggestion is that you should adopt a systematic approach to looking at each diagram. Always follow the same steps because otherwise you run the risk of missing something critical.
- The third bit of advice is to look at the surrounding years when you find something unusual. A year will be explained by the "company it keeps." If only one year follows a pattern then it is a fluke. But if the surrounding years have the same pattern, then you have found a trend.
- The final suggestion is evaluate what you see in these diagrams with what else you know about the changes in the company's operation. As an example, if your company is moving into or out of an area of exposure, do you see the expected changes from your previous normal pattern? [2]

Let's begin looking at the diagram.

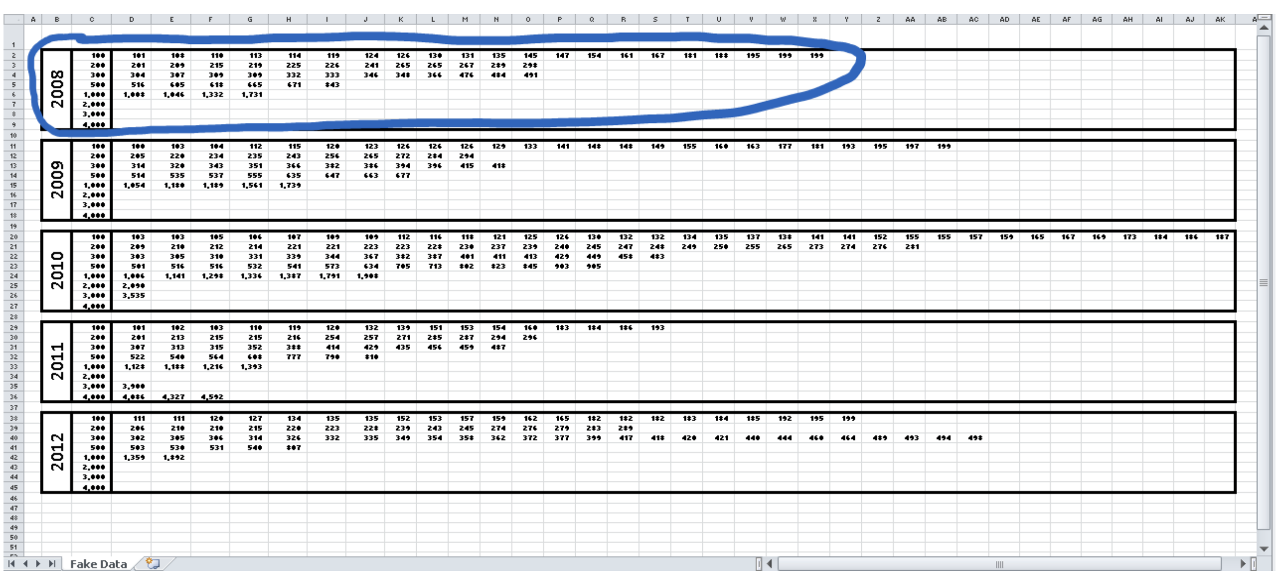


Figure 3

I will make believe that after years of looking at my fake data, I have come to recognize that 2008, the circled group in Figure 3, is a normal pattern for a year. It has a certain amount of claims in the lower layers and as we go up to higher layers, we see fewer and fewer claims until we get to the upper layers where we have no claims. If we look at 2009, the grouping right below the circled data, it is basically the same pattern. There are some random variations between the two years but basically we have the same pattern.

As far as a systematic approach to looking at the data, I am going to suggest looking at horizontal variations in the aggregate groupings and then focusing on individual years. If we look at all the years in the stem-and-leaf diagram, the first thing that catches our eyes in Figure 4 is that 2010 appears to be a bad year. You would already know that 2010 was a bad year from looking at the traditional development triangle but this diagram shows the additional insight that losses were coming in all layers as opposed to several large losses.

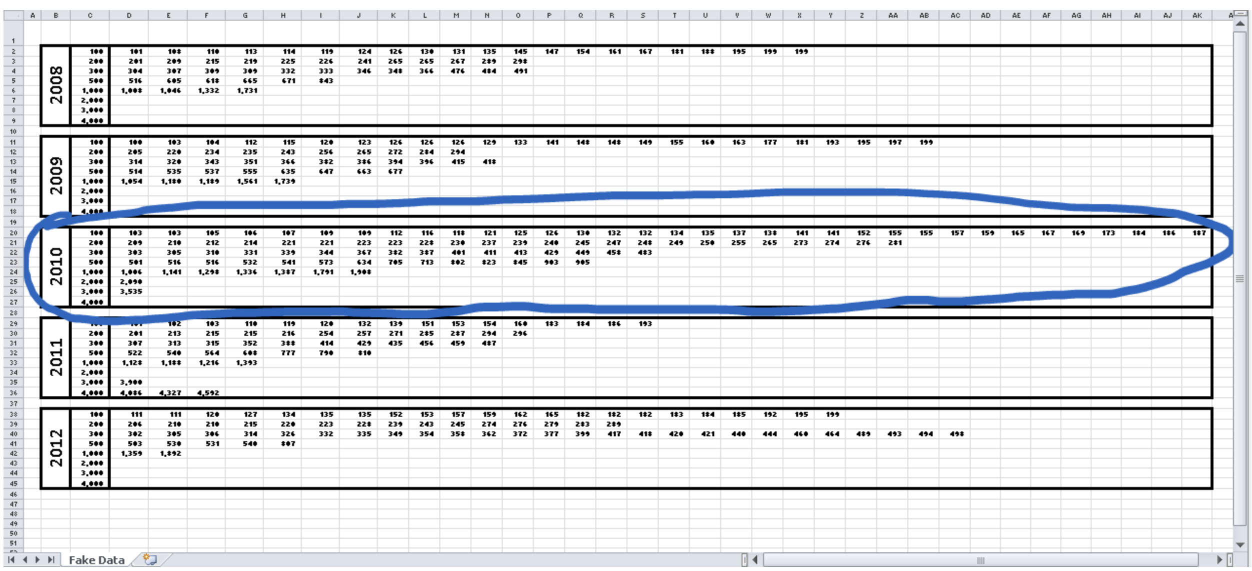


Figure 4

The 2010 stem has a lot of leaves on them. When looking at those lower layers, the leaves extend well beyond our normal year for all the layers. If you were trying to paint a picture of what was going on with a particular year, a diagram like this would help explain things. If we compare this year to all the other years, it looks unique. So this year was not part of a longer term trend.

Now let's look at some of the individual layers for the individual years. The year 2011, the circled year in Figure 5 on the next page, was the second worst year of them all. However, in this case a number of large losses are causing the problem as opposed to a frequency of losses. We can see an unusual number of large losses showing up in this year. In fact, you can quickly see that "this would have been the best year we ever had if it hadn't been for those large losses." I am going to go out on a limb and say that explanation probably does not appeal to you but at least you can see that it is true by looking at the diagram. The rest of the year looks better than our normal year. And this occurrence of large losses seems isolated to that one year. We do not have any evidence that we are moving to a new normal.

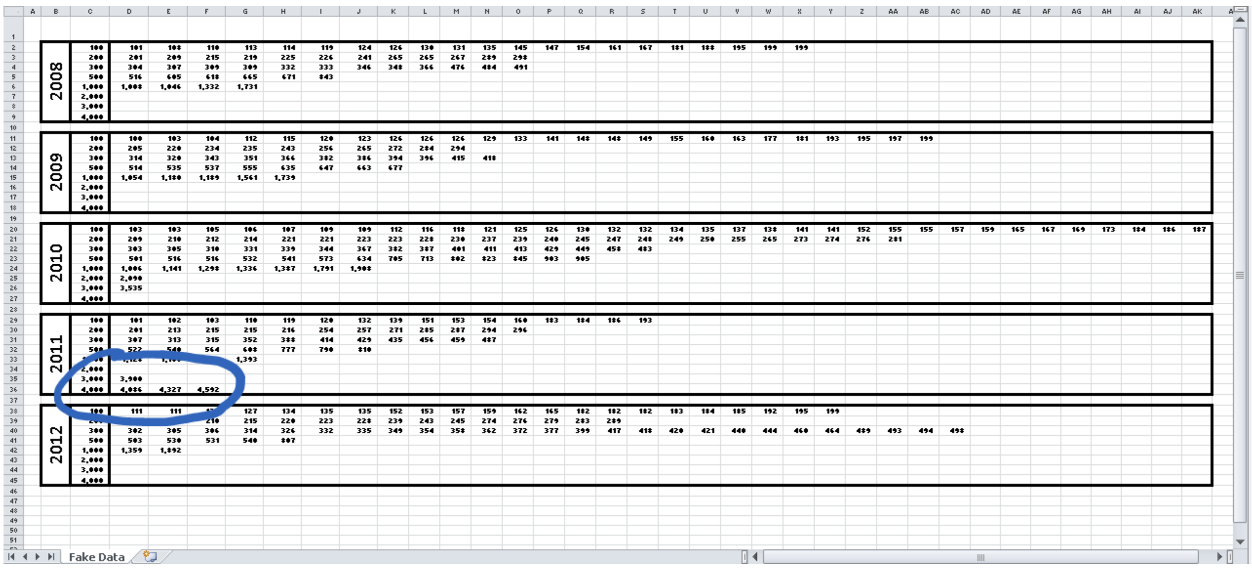


Figure 5

I will leave 2012 for you to look at and think about. How does it compare to a normal year? Using a systematic approach, do you see any unusual patterns as compared to our selected normal year? How does it compare to the prior year? Finally, you would want to ask how this year fits in with what you know about changes in the company's book of business. If the underlying exposure or type of business was changing, those changes would be part of your explanation.

Stem-and-leaf diagrams will help you quickly and clearly get your point across to the "big picture" people who are interested in your company's results. These diagrams may be a way to open up discussions with other interested parties. Just remember people may have alternative explanations for the observed changes than the explanations you offer. There will be different explanations for changes in the displays and that is what makes actuarial work and reading ECG's both an art and a science. And even though one of you might be barking up the wrong tree, hopefully these discussions will lead you in the right direction so that you all can get on the same page.

I started the paper with a 500 year old idiom. I will finish here with a relatively new one that may or may not be around in 500 years. Ladies and gentlemen, Elvis has left the building.

## REFERENCES

[1] Mason, Robert D., Lind, Douglas A., Marchal, William G., *Statistical Techniques in Business and Economics*, 10th Edition, pages 31-33.
[2] Garcia, Tomas, B., Holtz, Neil, E., *Introduction to 12-Lead ECG, The Art of Interpretation*.

### Biography

**Kirk G. Fleming** has worked as an actuary for over 30 years and has given many presentations to big picture folks. In addition, he has experience working in hospitals doing 12-lead EKG's.