



## Transcript for Session 038

Listen to the podcast session, see resources & links:

<http://chandoo.org/session38/>

### Transcript:

Hi and welcome back to <http://chandoo.org> podcast. This is session number 38 of <http://chandoo.org> podcast. Our podcast is designed to make you awesome in data analysis, charting, dashboards and VBA using Excel.

In this episode we are going to talk about **data to ink ratio**. It is a very popular term and it is a very useful technique (the concept denoted by this term) to creating powerful, amazing and insightful charts whether you are using Excel or any other software for that matter. So, we are going to talk about that technique. But, before we jump into the podcast topic, I just want to share a couple of announcements with you.

As you probably know, I am running a new course called 50 Ways to Analyze Data. We opened this course for enrolment in February 2015 and we had massive success - more than 350 students signed up for the program in February. I am **re-opening the doors** for this program - **50 Ways to Analyze Data** - on **Monday, 13th July 2015**. As you can see, it could be just around the time that you are listening to this podcast or maybe a little in the past. So, the date is 13th July 2015. The enrolment window opens on Monday, 13th July and it closes on Monday, 20th July. As you are a podcast listener, I would like to reward you with a small gift. Here it is - you will get a discount of \$38. That's because this is podcast session 38! Now you wish that I had been a little more prompt and recorded more episodes. But, anyhow, you will get a discount of \$38 which is roughly 10% of the course fees. To claim your discount please use the **discount code** "PLAY38". PLAY stands for 'Podcast Listeners are Yummy!' So, PLAY38 is the discount code. If you use this during the check-out process, you will be able to claim a discount of \$38 on the 50 Ways to Analyze Data program. Please visit <http://chandoo.org/session38/> for a link to the 50 Ways to Analyze Data training program where you can go and learn more about this training program and sign up for it.

Just in case you are curious, as per the name of the course "50 Ways to Analyze Data", it comprises of 50 case studies about how to do data analysis. The course is designed for Analysts, Business Analysts, Managers, Reporting Professionals, Business Intelligence Professionals, Data Scientists and such people who need to use data and analyze it but they don't know where to start or how to go with various kinds of analysis. It is not your run of the mill intermediate level course. It is a fairly advanced program. We



are going to talk about techniques ranging from statistical analysis, financial analysis and marketing analysis. Very advanced stuff is included as well like segmentation, cluster analysis, outlier analysis and all sorts of things like modelling, optimization problems and really in-depth and very elaborate programs. There are fifty case studies. Each case study deals with one type of business problem or analysis scenario and showcases various solutions and approaches and techniques based on formulas or pivot tables or Solver or statistical analysis tools or one of those kinds of things along with the presentation of that result to the management through output like interactive charts or presentation model or workbook modelling etc. It is a really well-designed course. I got lots of positive feedback from the students who attended the first batch so I am sure you are going to love it as well. Just keep in mind that you will get a \$38 discount. Your discount code is PLAY38 and you can go to <http://chandoo.org/session38/> for more details about this program. That's the first announcement.

The second announcement is that I am running a **webinar on how to be a better analyst**. This is the first time I am running a webinar on <http://chandoo.org>. The webinar will happen on **Wednesday, 15th July**. The course is launching on the 13th and I am running a webinar on the 15th. You are welcome to participate in it. It is a free webinar. Essentially, the webinar has two parts - the first part which would be roughly 30 minutes talks about how to be a better analyst and suggests a framework and concepts and ideas for improving your job prospects and work skills as an Analyst. The second part of the program talks a little bit about the 50 Ways to Analyze Data course and how that will help you in your journey to become a better analyst and offers some more details about the course and the benefits to you. It is on Wednesday, 15th July, 2015 and I am not really sure if you will be listening to this podcast before 15th July 2015. If you are listening to this well ahead in the future you can go to <http://chandoo.org/session38/> where you may be able to find a link for replaying that webinar. Either way, go to <http://chandoo.org/session38/> where you can find the link to both the webinar registration as well as the replay once the time is up. So, please go and enjoy the webinar as well as your special discount for being an awesome podcast listener for the 50 Ways program.

Those are a couple of announcements. Now that we are done with them, let's talk a little bit about data to ink ratio. What is this data to ink ratio? Data to ink ratio is a concept that I think I first came across through Edward Tufte who has authored many books and is considered to be a guru when it comes to creating better charts and visualizations. He proposed a lot of ideas and concepts that kind of stand the test of time and even today people use them to decide how to make a better chart and how to clean it up and how to format it. So, data to ink ratio is, I think, a concept that he proposed first. I may be wrong here because, based on the reference books that I have, it seems that he is the one who proposed it.

The idea is basically like this - when you print out a graphic whether it is a chart or visualization or an info-graphic - when you print it out on a paper, the amount of ink dedicated to the data part of the chart versus the total ink is what data to ink ratio refers to. This means that if the data to ink ratio is low then it means that the chart has too many things that are not really contributing to the data part of the chart. For example, if a chart has a 3-D shadow applied to it or some sort of a weird reflection effect applied to



it then the reflection is not going to help you understand the data better. It is just there to give a bit of glamour to the chart. So, it reduces your chart's data to ink ratio. This is what the data to ink ratio refers to.

Now, these concepts are proposed probably based on graphics or visualizations that appear in news media etc. But, nowadays, not many of us are printing things. So, you could go and consider data to ink ratio as data to pixel ratio because, essentially, how many pixels or screen space is taken up by a particular graph versus what is the total screen space or number of pixels dedicated to the actual data portion of the graph. For example, in a line chart, the line itself would be the data part. So, if there is only a line and nothing else then it has 100% data to ink ratio. So, this is the concept of data to ink ratio

Now, as an Analyst, how does knowing the definition of data to ink ratio help us? Where does this fit in to our workflow? Well, whenever you are designing a chart, you should aim for optimizing data to ink ratio within reasonable means. Why do I say within reasonable means? This is because, as I said earlier, let's take the example of a line chart or a column chart if only that line is present or only those columns are present and nothing else is there, i.e. no title, no axis, no legend and no gridlines. Then, technically, we have 100% data to ink or data to pixel ratio and it is optimized. We can't beyond 100%. But the thing is that the chart is also not very much readable or understandable. If the axis is missing we don't know what the range of that line is or where the columns stop or what the maximum column height is and what the minimum column height is. So, although technically we have optimized the data to ink ratio in such cases like a simple line or column alone and nothing else, we have now optimized it at the cost of understandability or readability of the chart. So, we need to optimize this within meaningful reason. That's what we should consider as Analysts.

Now, if we need to optimize data to ink ratio, what would that help us in? Well, when you optimize data to ink ratio, you end up with charts that are highly readable and highly presentable. This is exactly what we desire as Analysts, don't we? We want to create charts that are easy to read, easy to understand, easy to take home and implement the suggestions. If you are presenting a report to your boss, you won't want your boss to scratch her head and pull out all here hair before she can understand the charts. So, this is the intention behind data to ink ratio. Data to ink ratio as a conceptual framework will help us come up with charts that are better and easy to work with.

So, how do we optimize data to ink ratio. Well, there are a couple of obvious reasons and there are a couple of not so obvious ones. We will talk about them. For more about data to ink ratio, I suggest that you read books both by Edward Tufte and Stephen Few, especially Stephen Few's book 'Information Dashboard Design'. This is a great book and I referred to that book to prepare material for this podcast. So, I suggest referring to Stephen Few's 'Information Dashboard Design' book as a good guideline when it comes to designing beautiful charts, reports or dashboards.



Now that I have given you some reference material, let us talk about **ways to optimize data to ink ratio**. The obvious mechanisms and ways to optimize data to ink ratio would be that anytime you are using a three-dimensional chart like a 3-D pie chart or a 3-D column chart or a 3-D bar chart, the third dimension is not going to add any data. Essentially, all the pixels or screen space dedicated to the third dimension laying out that nice 3-D structure is doing zero to your data to ink ratio. So, removing 3-D effects is like the most obvious way to improve your data to ink ratio. So, once you get rid of that, your chart becomes instantly more readable. More space or more number of pixels of that chart are now dedicated to the data part. So, get rid of 3-D chart effects unless you must use them. This is the case - 'must use them'. When would I 'must' use a 3-D effect? Well, the only scenario that I can think of (to my sane mind) is if I have a boss and she asks to give her a 3-D chart or else she is going to fire me. Only then would I use it. Otherwise, I never deliberately create a 3-D chart. It has been a while since I created a 3-D chart. By a while I mean more than 5-6 years. I have never even clicked on the 3-D chart button in newer versions of Excel. Maybe I have used it in 2003, 2007 or maybe in 2010 but I can't remember using it in 2013 in the last several months or years. So, **get rid of the third dimension**. That's the most obvious way to improve your data to ink ratio.

The other obvious candidates for improving data to ink ratio is that gridlines, background colors, background shapes or background images on the chart or any kind of gradient effects, for example gradient filled columns and those kinds of things can be **filled with a single color or no color or made transparent or even be removed completely** like in the case of gridlines. When you do that, you are improving your data to ink ratio. Now, each time you delete something like gridlines or labels or anything like that like legend etc., obviously, you will be improving data to ink ratio but you would also sacrifice a little bit of readability. So, you have to **balance** this. Can I take this out and still keep the chart readable and presentable? So, you should think about these two things as an Analyst. There is a bit of give and take when it comes to data to ink ratio. So, those are the obvious candidates - your 3-D effects, background colors, gradients, shadows, reflections and all sorts of things that we tend to add either because we are enthusiastic or because we choose a template or theme that is enthusiastic and it is adding all those things. So, once you get rid of them, your data to ink ratio is improved.

Now let's talk about a few more techniques that are beyond obvious when it comes to data to ink ratio. One thing that stands out is that many people, knowingly or unknowingly, use separate colors for individual items in charts. For example, if you are making a column chart and there are six columns, instead of all the six being in the same color, you now have each of the six columns in a separate color. So, in my opinion, this is basically not going to help you with data to ink ratio. So, you can consider having all six in the same color so that the data to ink ratio improves.

Now, you might be thinking that the number of pixels dedicated to the columns is not going to change. The only thing that we have done is that instead of printing them in blue, green, yellow, orange, we now have everything in blue. How does that improve data to ink ratio? The thing to keep in mind is that although separate colors are not going to increase the number of pixels or the amount of ink given to



the chart, it is going to take more time for your users to process this. Each color is kind of creating a mental separation so the more colors that there are, the more time it takes for your users to understand them and appreciate these colors and put them in the context of the overall picture. So, in that sense, if you can get rid of the colors and make one color then it will improve your data to ink ratio. Now, this comes with a special note. The thing is that if the colors do mean something then you can't remove them. A good example is that if there is a clustered column chart, i.e. you have 6 columns with three columns together and then three columns together where the first three columns refer to product 1 and the second three columns refer to product 2 then in such a case, you are allowed to use different colors for the first three columns and the second three columns. So, red for the first three and blue for the second three is allowed. When you are talking about columns or bars or anything like that within a set of data then there is really no need to change or vary the color by data points. When you do that, you are reducing your data to ink ratio. So, to optimize it, change the color to the same and do it that way.

You might be thinking that this is fine but how does it work when I have a bunch of columns and I just want to highlight the column that has the maximum or minimum value. This is something that we do often. When you are looking at a bunch of columns or bars, we usually set up some sort of an extra dummy column or something like that, so that we can highlight the maximum value column or something like that. When you do it, you are obviously compromising a little data to ink ratio but I think this compromise is allowable because when you highlight the column with maximum height or the one with some sort of special meaning, it helps your readers understand your report better and make better decisions. So, **we need to balance the data to ink ratio**. We are not really striving for 100% data to ink ratio. We are striving for a happy medium of 70-80% or something like that where it gives a good balance and helps our readers understand the charts. So, **choosing one color instead of multiple colors is a non-obvious method**. Many people don't realise it but once you apply it, your chart will improve.

Likewise, many times when you have a dashboard, usually we have a lot charts in dashboards instead of one chart and people tend to add borders around individual charts so that it looks like there are lots of different boxes in the dashboard but maybe with a bit of clever formatting, you could get rid of borders and use the white space as a separation mechanism. So, elements like borders are also a good candidate for optimizing data to ink. Anything that is printed on the screen will get some attention from your users. So, hypothetically, if you imagine your screen to have one million pixels - if the screen is turned off then everything is blank but once you turn it on or print a report or dashboard there, some pixels will light up. So, for the sake of simplicity, assume anything that is not in white color will get our attention because, by default, everything in Excel is in white color. So, any pixel that is not white colored will get our attention. The most important ones that should get our attention are the charts where your colors and your lines and columns will go. But, because there are these borders and gridlines and other elements on the screen, they will also get our attention. So, we need to **minimize the attention that is going towards these non-data parts of the report** and borders are a good candidate for optimizing. You could get rid of them or you could replace the borders with a very thin and dull-colored grey line so that it doesn't get so much attention.



Any other elements that are usually found in a report or dashboard, for example, navigational elements like menus or links or buttons that you place are also non-data elements. So, we need to minimize their impact, amount of screen space or pixels dedicated to them. You could either completely get rid of them or, if that is not possible which is the case in many business situation, we could come up with mechanisms so that they are low impact and low key. So, change the colors, font sizes and their place on the screen so that they are positioned where we naturally give no importance. Usually when we look at an Excel screen, we tend to look at cell A1 first because that is the top left corner. So, your report should contain important elements in the top left corner or, in general, in the top area. All the data should go there. All the charts will go there and any other non-data stuff like your navigation buttons and menus etc. can be further down the report or towards the right hand side. This depends on the context. If you are talking about an Arab country where the usual reading is not left to right but right to left then in such a case, you need to reverse things and place the most important items on the right hand side of the screen.

Once you optimize your data to ink ratio, the next step would be that you need to **highlight what is important**. The concept of data to ink ratio is that we want to de-emphasize, i.e. get rid of things that are not important. At the same time, we also need to emphasize things that are important so that your report, all in all, looks meaningful and anybody looking at it will know which pixels or which area or which ink is more relevant and more important for their business decisions. So, to do that, you could use concepts like different font sizes, different colors or even using conditional formatting to highlight certain parts of the report.

Once you follow all these steps, you would then have a report, chart or dashboard that has a very high data to ink ratio. Keep in mind that we are not aiming for 100% data to ink ratio. We are aiming for something that is a good balance between data to ink ratio and showing a report that is also readable.

So, there are many good examples of data to ink ratio. Please visit <http://chandoo.org/session38/> where I am going to point to a couple of resources, books and examples that have a very high data to ink ratio and how that helps you in terms of coming up with a report that is really good looking and informative. So, please visit <http://chandoo.org/session38/> where you will find all the show notes, resources and links mentioned in this podcast. Just as a reminder before we wrap up, I want to tell you about the 50 Ways to Analyze Data program that is opening up for enrolment on Monday, 13th July, 2015. As a podcast listener, you get a \$38 discount on this program. Please use your discount code PLAY38. PLAY stands for 'Podcast Listeners are Yummy'! PLAY38 is your discount code and I hope to see you in our 50 Ways program so that you can be an awesome Analyst. Thank you so much. Enjoy your day. Bye.