Chapter 6: Improving your Visualizations

“The human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest bandwidth channel into human cognitive centers.”

Colin Ware, “Information Visualization”

In chapter 1, we briefly mentioned how data analysis is like how we imagine archeology to be: spending hour after hour with small tools in the hope of uncovering even the tiniest of insights in the earth. That analogy can be extended into the shared desire to create a narrative. Archeologists attempt to recreate the stories of history by digging up parts of a story and it’s the same with data analysts. There are stories buried in the data and it’s up to the data analyst to uncover that narrative, piece it back together and communicate that story to others. The data always should be communicated in its context and setting. We’ll want to identify the characters and walk through the relationships and events in the data. The reader should be able to follow the story as it unfolds right through until all questions and inconsistences are resolved. Thankfully as we dig around in the data, one or more narratives will naturally form and then we have the challenge: how can we best communicate the stories within our data to others? The answer is within the field of data visualization and is the focus of this chapter.

Our ability to visually process information is by far the most efficient path to human understanding. Like a good hacker, we want to learn about this system, understand how it functions (and how it doesn’t function) and then exploit this cognitive system to achieve our goal. In this case, our goal here is effectively and efficiently communicating the stories we find in our data.

Research into the science of perception has exploded in the past few decades. Psychologists, neuroscientists and others are chipping away at the mysteries of the human brain and slowly, the rules of how we visually process information have been emerging. Understanding the rules around how the human brain visually process information will be important in our reports, presentations and dashboards. But even more than that, we’ll want to also use visualizations in our analysis, and create a communication path from the data to us.

Why Visualize?

There are many advantages to using data visualization as a communication tool compared to other methods. To paraphrase Colin Ware (who we quoted to open this chapter), data visualization has the following advantages:

* **Data visualizations communicate complexity quickly.** Descriptive statistics (mean, median, variance, etc.) exist to describe and simplify data but tend to remove subtleties that may exist in the data. By visualizing the data, it’s possible to communicate millions of data points in seconds while minimizing the loss of detail and resolution.
* **Data visualizations enable recognition of dormant patterns.** Often times, visualizing data enables us to see patterns that would never be apparent using statistical methods or scanning the data. By visually representing the data, often times the patterns in a single variable or relationships across many variables may leap off the screen at us.
* **Data visualizations enable quality control on our data.** By visualizing the data, often times mistakes and errors with data collection or preparation become apparent. Data visualizations can serve as a good and quick sanity check on our work.
* **Data visualizations can serve as a muse**. It’s been said that most breakthroughs in science didn’t start with a “Eureka” but instead with a “Huh, that’s odd.” Laying out our data visually can give us new perspective and help facilitate our thinking and discovery process.

The fact that we’re focusing this chapter on visualization does not mean that visualizations are always the best way to communicate data. If the analysis can be summed up with a sentence in an email, or perhaps a simple table in a report, there’s no reason to force it into a visualization. Our focus is on the successful communication of the narrative; the method of communication is just a means to that end.

Unraveling Visual Perception

The entire system and process of we process of visual information is incredibly complex and much of our knowledge around it is still evolving. However, there are a few key (and hopefully easy) concepts that we should understand because how the brain visually processes information will help us create great visuals, or in some cases it will help us understand how not to create visuals.

We begin this journey with visual stimulus in the form of light that our eyes convert into electrical signals for our brain. This information will pass through stages of our **visual memory,** each with a specific set of strengths, limitations and functions. Before we are consciously aware of it, our brains rapidly scan the visual field, which is called **preattentive processing**. Finally the brain will instruct the eyes to focus elsewhere, and through a series of **saccadic movements** our eyes will focus on various features to help build up the image in our mind. With these three aspects of our visual processing system, we should be able create a solid foundation for good visuals and dashboards.

Visual Thinking

**Iconic memory** is the first stop for the visual information. It is a very brief stop, lasting around half a second or until it’s replaced with new information. But what happens in this tiny window is critical to creating good visualizations and dashboards. With the information stored in iconic memory, the brain preprocesses the image prior to giving it any conscious attention. From an evolutionary perspective this is quite helpful, this preattentive processing can help us quickly identify possible threats in our environment. For example, anyone who has been driving when an animal dashes in front of the car has probably felt that urgent message from the brain when it recognizes a possible threat. We begin to react immediately even before we can process the full extent of the threat. While we hope our visualizations aren’t treated like a threat, it’s that visual searching and preattentive processing that we can leverage to draw attention and even communicate some basic attributes of our data to make processing much easier when we begin to consciously process it.

**Working memory** is the next stop and things get a little more complicated here. First the brain will gather up and group visual aspects into meaningful objects and hold these individually in working memory. There is a lot of flexibility within working memory as we can rapidly replace or drop these objects as we take in more information, but the flexibility comes at a cost in capacity. We can only hold three to five objects in working memory depending on the task and objects. This limit is important when designing visualizations and dashboards. If we create a visualization with a legend that has ten different attributes, the reader will have to continually reference the legend in order to understand what they’re looking at. So as we communicate the stories in our data we want to limit each visual to no more than five objects (four to be safe).

**Long-term memory** is not directly important as we attempt to communicate our data other than both iconic and working memory are temporary stores. In order for something to move into long-term memory the reader needs to visually “rehearse” the information to transition that visual chunk from working memory into long-term memory. But indirectly, we will leverage long-term memory to detect meaningful patterns and relationships within the data. This type of deeper understanding and processing is only available with long-term memory.

Tracking Eye Movements

When we focus on something like a dashboard and/or computer screen, we do not simply fix our gaze on it and take in the image as a whole. In reality, our eyes will dash around the screen, focusing on very small portions for very short periods of time in order to build up the image in our mind. One eye movement is called a saccade, overall they are called saccadic movements and they are not random. The brain has a (rather confusing) set of rules (guidelines really) for how the next fixation point is prioritized. For example, when another person greets us, our eyes perform scanning saccades over their entire face, bouncing from the distinct features of the face (eyes, nose and mouth) and establishing the edges. The same applies to our visualizations and dashboards. The eyes will fixate on an obvious feature and bounce around and between to the points it considers important in order to build up the entire picture.

The saccadic motion itself is largely unconscious and is thought to be a ballistic movement. Meaning once the brain initiates a saccadic movement, the muscles take over and handle the rapid acceleration and deceleration from beginning to end. This is important for two reasons: once it is initiated it cannot be changed or stopped and during the motion we suppress much of the visual input.

All of these points can be rolled up into a few important learning points from saccadic eye movements. Knowing that the eyes will bounce around from feature to feature and the ballistic nature of the movement, we should keep several points in mind as we create our dashboards and graphics:

* **Don’t overload the dashboard with visual features**. Keep the number of attention-grabbing features under control because if everything is important visually, than nothing will be important visually and the analyst will have to put more effort in to understand the visual.
* **Make the important messages obvious visual features.** Just as we will scan the important parts of a human face, we will look for the similar attention-grabbing features on the screen. Make sure that those features are clear and are important to the viewer.
* **Limit time wasted on saccadic movements.**  Saccadic movements that jump longer distances take longer to execute. Do not push the visual features into the corners or towards the edges. Forcing the viewer to bounce across large distances will decrease the amount of time they are actually seeing the features (and increase the time spent in saccadic movements).

The role of saccadic movements influence dashboards much more than static data visualizations. Typically in a static visualization we will have one, perhaps sometimes two visual features we want draw attention to and the eye movements are contained in a relatively compact space. But in a dashboard we may be trying to communicate several data points simultaneously with varying degrees of urgency. Good dashboard design, as we’ll cover in chapter 10, will want to limit the time spent in a saccadic movement and exploit the eye movement for efficiency in our communications.

(when to use tables is few’s book)

The preattentive processing detects several attributes, such as color and the location of objects in a 2-d space. Because preattentive processing is tuned to these attributes they jump out at us and are therefor extremely powerful aspects of visual perception. If you want something to stand out in a table or graph, you should encode it using a preattentive attribute that contrasts with the surrounding information such as red text in the midst of black text. If you want things to be seen as a group, assign them the same preattentive attribute.

ther

Preattentive Processing

The best way to describe preattentive processing is through pictures. Take a look at figure 6.1 and try to count how many capital X’s are in this completely random mix of letters and numbers.

Figure 6.1 Count the number of “X” characters [FILENAME 793725c06f001]

Because all of the letters are the same color and contain the same relative space, nothing stands out. The brain simply sees a collection , in order to count the X’s we have to scan through each letter across the four rows. While we’re doing that we have to remember how many we’ve found as we scan so we don’t lose track. But now take a look at this completely random mix of letters and numbers with the X characters emphasized so we can easily pick out the character.

Figure 6.2 Count the number of “X” characters [FILENAME 793725c06f002]

Now when we fix our gaze on it, the brain sees a background of gray symbols with 4 similar objects. The brain will mentally create two groups, one group of all the gray symbols and a second group for the dark red X’s. When we consciously recognize the second group as what we’re interested in (the X’s), it becomes trivial to visually exclude the gray characters and now we can scan just through this group. Counting the X’s becomes a simple and quick task.

That mental grouping and ease of focus is what we are after. We want to enable our preattentive processing to effortlessly group similar objects and highlight where we want attention to be focused. But we have to keep in mind that the preattentive processing is not all that smart. There are only a handful of attributes that our preattentive processing will be able to pull out. The sole purpose is to recognize features in our visual environment. It will not be able to project meaning, interpret the objects or make meaningful associations (beyond simple visual grouping).

Through literally hundreds of studies, researchers have been able to differentiate visual attributes from what can be preattentively identified from those that can’t. Having looked through some of these studies they can get a little silly and abstract (how easy is *parallel* detected?), but looking at them as whole, we can create some high level categories of what can be preattentively processed. These categories are form (line, shape, size), color (hue and intensity), spatial position (two-dimensional, stereoscopic) and motion (blink, direction). And the list of specifics can get quite long, but thankfully we can experiment here. Feel free to create some visualizations and try out different ways to highlight the data. Chances are good if it’s easy for you to pick out, it’ll be easy for others (and it’s a good idea to run things by others as a sanity check). Figure 6.3 gives a few visual examples of ways to differentiate based on preattentive attributes.

Figure 6.3 Examples of Preattentive Attributes [FILENAME 793725c06f003]

Not all preattentive attributes are created equal. Look at figure 6.3 again. While they all highlight the three data points, some make the three points slightly easier to see than others. For example in figure 6.3(e), if we would have chosen pink to show with red it would have been slightly more difficult to pick out the differences. The amount of “pop” for preattententive attributes depends on how different the attributes are. The shapes in example in 6.3(a) are more different from each other than the circles and squares in 6.3(b) and slightly easier to see. It’s still possible to see the difference in 6.3(b), but it’s just not as quick to “pop”.

This concept of preattentive processing should be treated as just that -- a concept. The line between our preattentive processing and conscious processing is gray and blurry. When looking at a visualization, we may slip between the two quickly and quietly. Some things considered preattentive, have been shown to be a conscious and learned skill depending on environment and culture. But the point remains for our visualizations and dashboards. If we want to direct the reader’s focus and attention we should leverage some basic elements like form and color to highlight the point we need to make in the data.

Finally, one last word of caution about preattentive processing: it’s possible to overload this process and negate any benefit. Take a look at Figure 6.4 below. In the first example 6.3(a), we separate three groups by color and it’s quite easy to pick them apart, not only are they spatially grouped, but the color highlights the difference. In 6.3(b), we attempt to communicate a difference with shapes and it’s a little harder to tell them apart, but we can still pick out the two groups. When we combine them in 6.3(c) things get a bit more complicated. Now to separate based on shape we have to actively inspect the elements and separate them consciously. We have to be careful to keep the visuals as simple as possible to exploit the readers preattentive processing.