Five
Simple
Steps

A Practical Guide to

Designing with Data

by Brian Suda

A Practical Guide to Designing with Data
by Brian Suda

Published in 2010 by Five Simple Steps
Studio Two, The Coach House
Stanwell Road
Penarth
CF64 3EU
United Kingdom

On the web: *www.fivesimplesteps.com*
and: *www.designingwithdata.com*
Please send errors to *errata@fivesimplesteps.com*

Publisher: Five Simple Steps
Editor: Owen Gregory
Production Editor: Emma Boulton
Art Director: Mark Boulton
Designer: Nick Boulton

ISBN: 978-1-907828-00-3

A catalogue record of this book is available from the British Library.

Foreword

Jeremy Keith

I have known Brian Suda for many years. We first met through the microformats community, where he uses his skills to make structured data readily available and easily understandable. Now he is applying those skills to the world of data.

I am covetous of Brian's mind. It is the mind of a scientist, constantly asking questions: "What was the first man-made object with a unique identifier?", "What would a hypercube of bread classification look like?", "What would it sound like to say all fifty of the United States at the same time?"

Okay, that last one was from Family Guy. But whereas you or I would be content to laugh at the joke and move on, Brian actually tried it by layering fifty audio recordings on top of one another. For the record, it sounds like this: Mwashomomakota.

As you would expect from such an enquiring mind, this book is not a shallow overview of graphs and charts. If you are looking for a quick fix on how to make your PowerPoint presentations pop, this isn't the book for you. But if you want to understand what happens when the human brain interacts with representations of data, you have hit the motherlode.

It isn't hyperbole to say that this book will change the way you look at the world. In the same way that typography geeks can't help but notice the good and bad points of lettering in everyday life, you're going to start spotting data design all around you.

Better still, you are going to learn how to apply that deep knowledge to your own work. You will begin asking questions of yourself: "Am I communicating data honestly and effectively?", "What is the cognitive overhead of the information I am presenting?"

Your mind will be more Suda-like once you have read this book. The phrase "change your mind" is usually used to mean "reverse a decision". I want to use the phrase in a different, more literal way.

This book will change your mind.

Introduction

Over the years, I have been digging through large data sets both for work and pleasure. I love numbers, charts, graphs, visualizations, zeitgeists, raumzeitgeists, infographics and old maps. Getting to peek into what companies like Google get to see on a daily basis – trends, fads, search volume, relatedness, all bundled up in an interesting illustration – makes my day. Some people re-read the same book over and over; I can stare at a dense illustration and re-read its story. It makes me ask, "What caused these numbers? Where did they all come from?" It has been estimated that the Large Hadron Collider produces fifteen petabytes (fifteen million gigabytes) of data a year. It's impossible to look at a table of fifteen petabytes of information – there has to be a graphical representation for anyone to comprehend data at this volume.

This is what excites me: the challenge of how to take these boring numbers and design something more compelling. To tell the story behind the data, we need to stop grasping for the perfect visualization and instead return to the basic language of charts and graphs. Only then can we begin to uncover the meaning and relationships the data has to offer.

Beyond the basic bar charts and line graphs taught in schools, a new breed of illustrations has recently appeared. These new 'visualizations' are an attempt to explain the underlying information with a powerful visual impact. They take complex ideas and distil them into beautiful graphics revealing the interrelationships in the data. Some are so brilliantly executed that there are now annual awards for newspaper and magazine infographics to highlight their achievements. Sadly, over recent years terms such as visualization and infographic have been bandied around with almost no regard to their proper use or meaning. Existing chart types and even slide shows have been saddled with the more gratuitous term 'infographs' to sound more impressive. There is a new vernacular in the realm of data representation, but that doesn't mean we should ignore the underlying principles and best practices of humble charts and graphs. Once you have mastered the basics, more complex designs and visualizations become easier to create.

I wrote this book because I feel that people aren't taking the fundamentals of graphs and charts seriously. Many people are inspired by fancy visualizations and jump right in over their heads. As with any discipline, you need to put in the hard work by starting from the beginning.

If you look at publishers' catalogues, there are plenty of books on this topic, but they all cover somewhat esoteric aspects of specific charts and graphs: either from an academic point of view, stating how to right align numeric values or calculate confidence intervals to two standard deviations; or illustrated guides to beautiful posters and the information they represent. While these topics are certainly important, you need to consider the data's context and the readers who are looking to your charts and graphs for answers. Having a beautiful poster or a well documented confidence interval is worthless if the rest of the design is unreadable. I wrote this book in an attempt to distil as much knowledge as possible into just the information you will need day-to-day. There are plenty of specific kinds of charts for very specialized fields, from financial to weather data and everything in between, but they all require a fundamental understanding of the basics. Just because someone might be an expert in their field doesn't mean that they have the know-how to design with data.

This book is a peek into the pinball machine of my mind, always bouncing around various related and sometimes unrelated topics. I wanted to draw together several techniques you can use in your charts and graphs, such as how to minimize the number of pixels, and at the same time explain some interesting aspects of colour in our lives. In addition, I felt it was important to explain how to spot bad or misleading design: the kind that unscrupulous people use to trick us into believing their interpretation of data rather than the facts. Only then can you properly focus on the data, bypass unnecessary distractions and avoid misrepresenting the information.

The main purpose of this book is to encourage you to visualize and design for data in such a way that it engages the reader and tells a story rather than just being flashy, cluttered and confusing. Much like authors choose their words, sentences and paragraphs to structure ideas, informaticians have a bevy of concepts at their disposal to tell stories with data. What I want to focus on are best practices. These pages don't describe the one true way to design with data, but set out some guiding principles you can have at your disposal when you sit down and illustrate data.

The Five Simple Steps team has been great during this whole process. They let me write the book I wanted to write and were incredibly patient with my digressions. I hope that some of them help explain my thought processes as well as entertain with some interesting tidbits now and then. The five in Five Simple Steps is a great constraint for a writer, but it does mean that this is hardly a complete list of types of graphs and charts. New ideas and concepts are always appearing, some good, some bad. Instead of following an endless trail, this is an overview of the common types you'll run into on a regular basis and how to push them to get the maximum value.

The title *Designing with data* was chosen because I wanted to focus not on data collection, statistics or other mathematical analysis, but rather on the visualization of data. Much like you create a narrative in words, illustrations of data need to tell a story. How you go about designing with data is just about as open as how to write a book. The one thing we do know is that it is important to get the basics right. This book is an introduction to those basics and an invitation to you to become the next expert data designer.

Contents

Part 3

Part 4

Part 5

Part 1

The visual language of data

Whether writing words, sketching a cartoon or illustrating a graph, you are telling a story to the reader. We all know the emotional power of a good book. The words on paper allow you to get sucked into the characters and plot. Classic paintings do the same. Millions of people go to see the Mona Lisa every year. Her eyes, her smile tell us a mysterious story. Who is she? The oil painting draws us into another world that we are welcome to explore. An effective illustration must do the same.

This book is about designing with data. Charts, graphs and other data visualizations have a language of their own. They convey meaning and information that is not available in words while demonstrating relationships within the data, and they allow the reader to make projections and better grasp the concepts. We need to learn how to tell a story with data and how to design it in such a way that it is no different than a great work of art or a bedtime story you remember from childhood. Well designed data should provoke emotions, tell a story, draw the reader in and let them explore.

Graph genesis

Chart literacy

Dynamic and static charts

Does this make me look fat?

Chart junk

1

Graph genesis

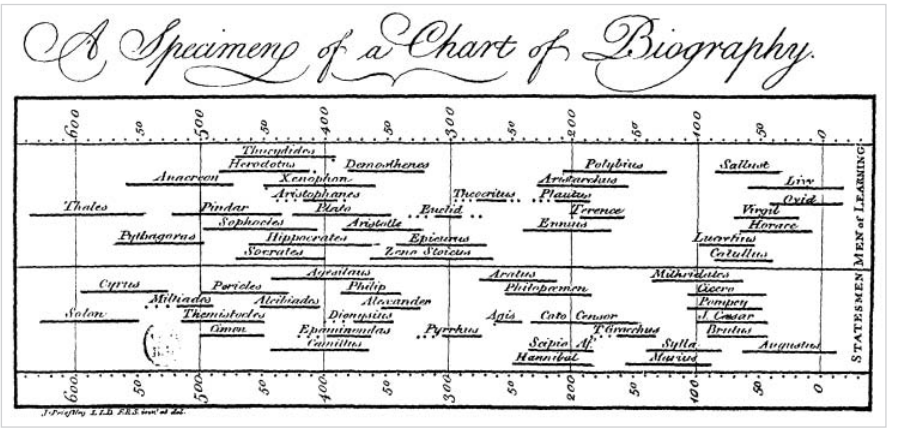
Some of the objects we use in our daily lives are so ubiquitous that we assume they have been around for all time. Even something as simple as the humble directional arrow, pointing the way and giving us instructions, once didn't exist. It was 'invented' for its modern day use. The barcode is a very modern invention (first conceived in the late 1940s but not widely used until the 1970s), yet we see it everywhere. One day, children will never have known a world without e-mail, the Internet or 3-D projection televisions.

In this book, I address the common types of charts and graphs, discuss some uncommon types and the reasons why they should stay that way. It's also important to look back at how many of these creations started off. As we'll see, they weren't so different than the charts we use today – two hundred years on and not much has changed. It makes you wonder, were charts and graphs invented or simply discovered? Were they destined to arrive in the form they did or are they so useful and near-perfect that we have just stuck with them?

Believe it or not, there once was a time before graphs and charts – they too were invented. The ancient Egyptians didn't have PowerPoint® presentations or pie charts (probably they also lived in blissful ignorance of a world of bulleted lists). Socrates taught his followers without the use of bar charts or Venn diagrams. Even the Romans didn't use graphs to visualize their luxurious spending sprees over the years, or miles of new roads constructed throughout the empire. Yet we assume that we can't live without three-dimensional quarterly projections from Excel®.

If we look back to the time when some of the very first charts were created, we'll see that they were born out of the need to explain large amounts of financial, political and social data. It wasn't until William Playfair (1759–1823), a Scottish engineer and political economist, published *The Commercial and Political Atlas* in 1786 and *Statistical Breviary* in 1801 that many of our modern visual devices for data first saw the light of day. He is credited with inventing the bar chart, line chart, pie chart and circle graph (see part 4).

Though Playfair gets most of the credit for kicking off this revolution, there are several other notable contributors who developed the tools we use today. You might know Joseph Priestley as 'the inventor of air'. Priestley was born in 1733 in England where he spent his life until 1791, when he fled persecution to the newly established USA. In 1765, he published *A Chart of Biography* which outlined the years of births and deaths of "statesmen of learning". This is one of the first timelines showing relative length and durations as a chart rather than as a table.



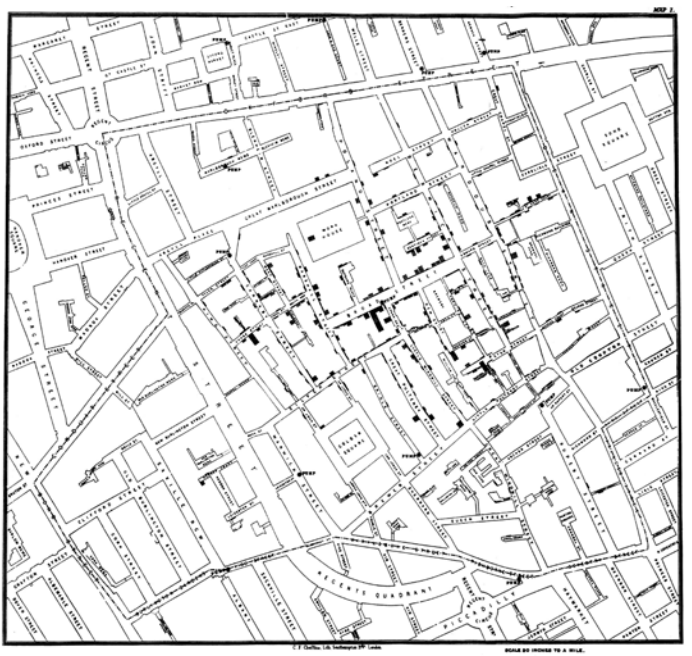
About twenty years after Priestley, in 1786, William Playfair published some of the first extant visual charts and graphs in his books about economics. Priestley's timeline certainly had an effect on Playfair as he created the bar charts. It is likely that William Playfair was aware of Priestley through various channels: the publication of Priestley's papers and books, almost certainly, but possibly also through the University of Edinburgh, which conferred the degree of Doctor of Law on Priestley in 1764. The Playfair family designed some of the most notable buildings of Edinburgh and its university, along with being deeply involved in the Enlightenment movement, as was Priestley.

Another contribution to the visualization of large sets of data came from Charles Dupin (1784–1873), a French mathematician and engineer. His contribution to the advancement of charts was

the choropleth map, first published in 1826. A choropleth map (sometimes called a cartogram – see chapter 21) is a diagrammatic map whose regions are variously coloured or shaded to illustrate a particular statistic, such as population density or voting choices. The first choropleth map designed by Dupin illustrated the illiteracy rate in France's regions.

A few years later, André-Michel Guerry (1802–1866) took Dupin's idea and began to build, embrace and extend. In 1833, he published his *Essay on moral statistics of France* which presented suicide and crime statistics in France on maps. The statistics were broken down and the maps coloured by just about every variable available. Guerry was fascinated by these charts and their social and economic impacts. He was the first to work in this field which later became known as moral statistics.

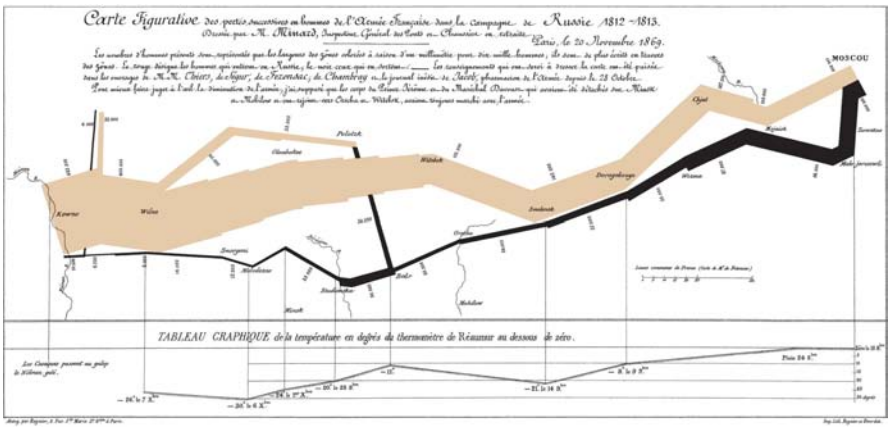
In 1854, a cholera epidemic swept through the city of London killing thousands of people every day. The health inspectors at the time believed that cholera was transmitted by infected air, but physician John Snow (1813–1858) believed that the outbreak was not airborne, but passed through the water supply. Using a map, he went door-to-door asking local residents about cholera deaths and marked each location to show the outbreak density. Using this data he traced the source of contamination to a local water pump and ordered the pump's handle removed, thereby preventing further spread of the disease in the neighbourhood. His map wasn't exactly a choropleth and it wasn't exactly a bar chart either. He mixed two chart methods to form a different kind of early visualization. Using geographically specific data, he concluded that the density of cases could be attributed to a local pump, strengthening his waterborne argument.



The next big advance in designing with data comes from Florence Nightingale (1820–1910). In 1858, while working as a nurse during the Crimean War, she emphasized the sanitary conditions in which the wounded patients were being treated. And there being multiple reasons why soldiers die while in hospital, she invented the polar area chart (see chapter 22), based on William Playfair's earlier pie charts from 1801. Her polar chart was a circular diagram of causes of death due to disease, wounds and preventable ailments. This illustrated the volume of lives that could have been saved given proper sanitary conditions: an early instance of hard statistical evidence being used to support changes in medical decisions and procedure.

In that same year, a French civil engineer named Charles Joseph Minard (1781–1870) pushed the boundaries of visualizations. He mixed multiple types of charts to convey more data and relationships. Taking a map of France, he overlaid several

pie charts to show the geographic positions of the data sets. A few years later in 1869, he published his now famous chart depicting Napoleon's march into Russia. This has been hailed as one of the greatest landmark works in data visualization history. It is a flow graph which uses a map to show the relative transfers of goods and people, geographic positions over time and several other related pieces of data such as temperature and troop size.



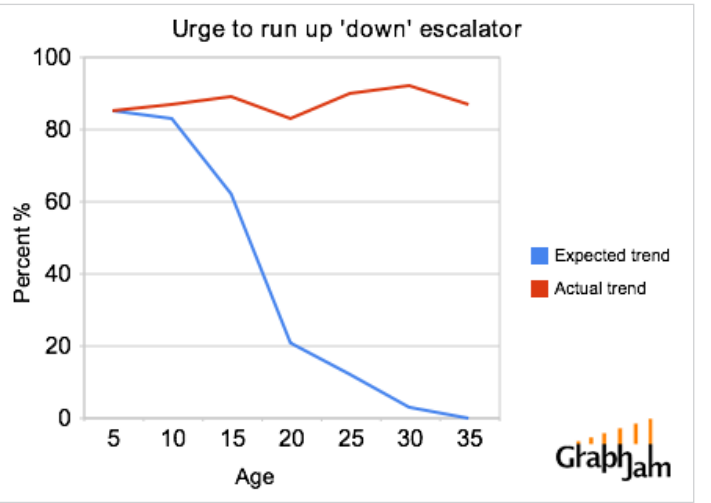
Within the last ten to twenty years, several new forms of charts have appeared. I attribute some of these to recent advances and innovations, but it's more likely that new charts and graph types are always being invented. Sometimes their usefulness is minimal and they disappear and no one remembers them. Only the recent developments that have not yet been proven useful or a burden seem innovative to us. Some of these new designs, such as treemaps (see chapter 18), will be discussed in this book because it is important to know their strengths and weaknesses before you jump in and use them in your own work.

2

Chart literacy

The amount of information rendered in a single financial graph is easily equivalent to thousands of words of text or a page-sized table of raw values. A graph illustrates so many characteristics of data in a much smaller space than any other means. Charts also allow us to tell a story in a quick and easy way that words cannot.

Graphs and charts are appearing more and more in our popular culture. Sites like Graph Jam[1] and Indexed[2] take concepts, song lyrics, observations and allegories and render them in rather silly or humorous pie and bar charts.



http://www.graphjam.com

We immediately understand them and they tell a story in their own right, sometimes more so than the original data. The fact that they give another dimension and life to the data demonstrates their worth.

Much of the value of graphs and charts comes from their clear, usable, legible interface whether on paper, screen or other medium. But graphs alone do not make something that is long, difficult and tedious instantly useful. It takes skill to be able to make a chart exciting and engaging without obscuring the facts.

[1] graphjam.com

[2] thisisindexed.com