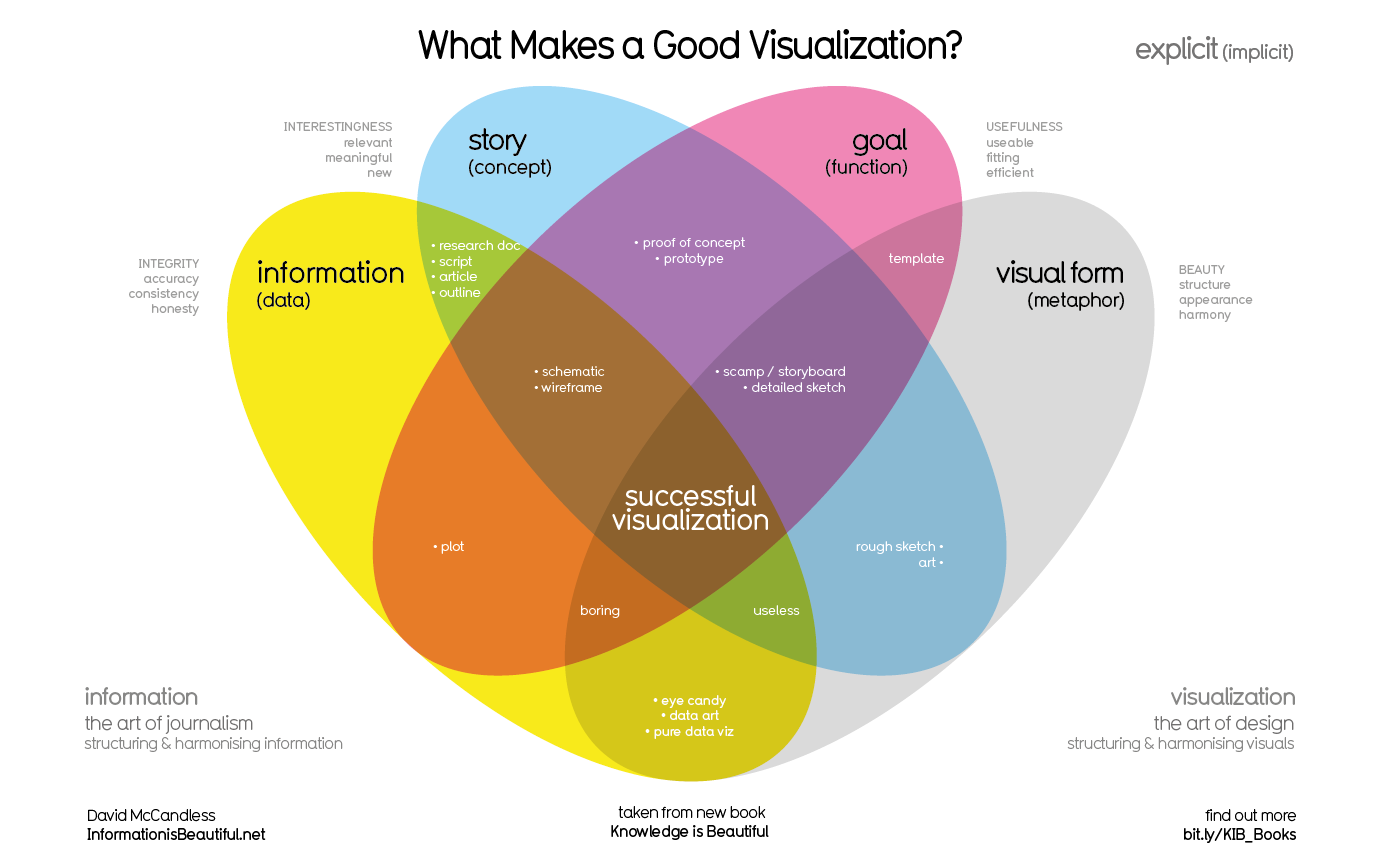
**Google Data Analytics**

## **Share Data Through the Art of Visualization**



[**The McCandless Method**](https://www.informationisbeautiful.net/visualizations/what-makes-a-good-data-visualization/)

1. **Information:** the data you are working with
2. **Story:** a clear and compelling narrative or concept
3. **Goal:** a specific objective or function for the visual
4. **Visual form:** an effective use of metaphor or visual expression

**Pre-attentive attributes:**

Pre-attentive attributes are the elements of a data visualization that people recognize automatically without conscious effort

### 

### **Marks**

Marks are basic visual objects like points, lines, and shapes. Every mark can be broken down into four qualities:

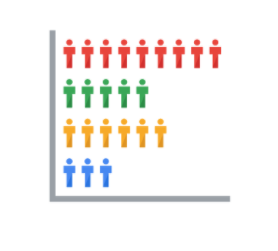
1. **Position** - Where a specific mark is in space in relation to a scale or to other marks



1. **Size** - How big, small, long, or tall a mark is



1. **Shape** - Whether a specific object is given a shape that communicates something about it



1. **Color** - What color the mark is



### **Channels**

Channels are visual aspects or variables that represent characteristics of the data

1. **Accuracy** - Are the channels helpful in accurately estimating the values being represented

2. **Popout** - How easy is it to distinguish certain values from others?

3. **Grouping** - How good is a channel at communicating groups that exist in the data?

**Design Principles**

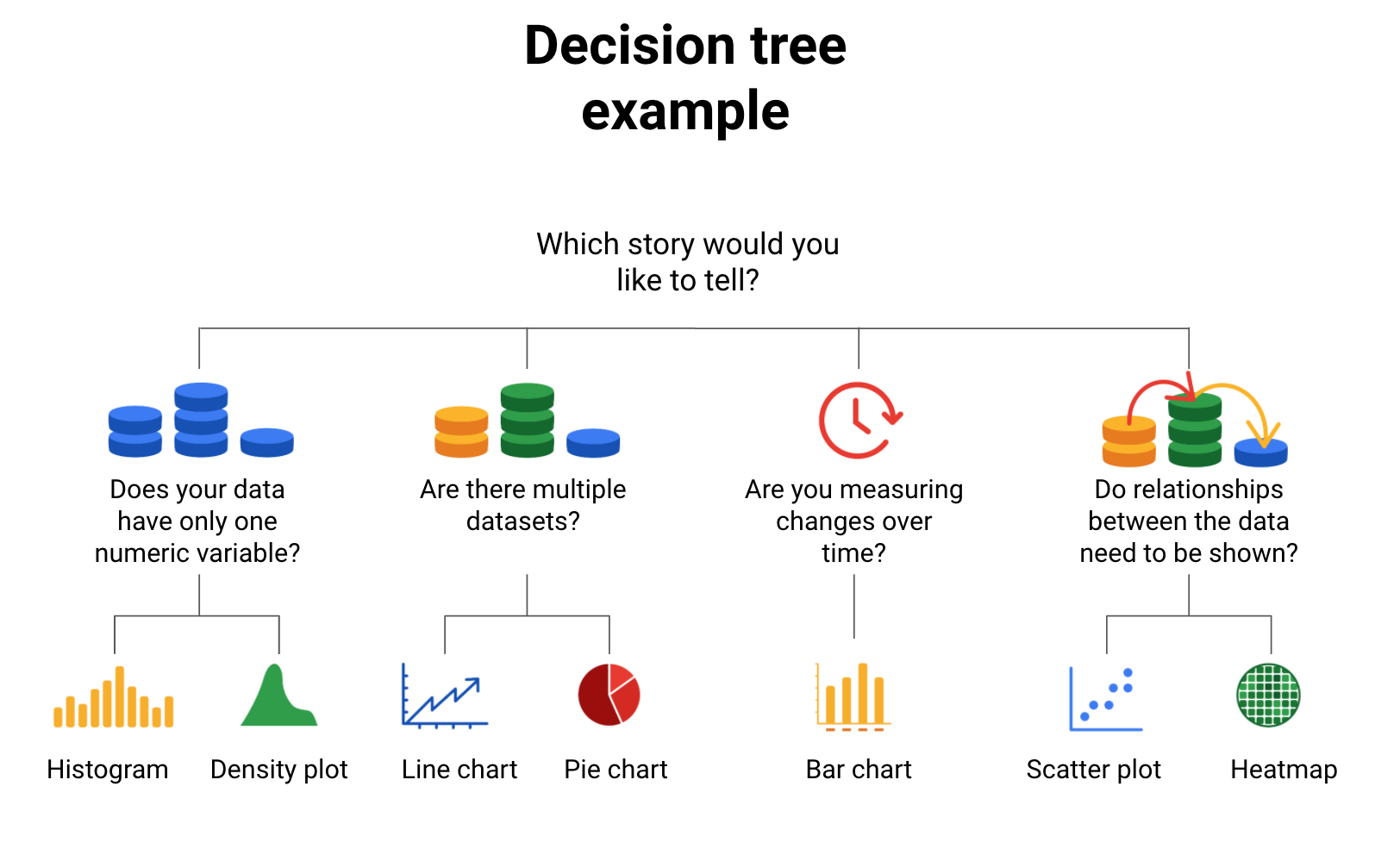
| **Principle** | **Description** |
| --- | --- |
| Choose the right visual | One of the first things you have to decide is which visual will be the most effective for your audience. Sometimes, a simple table is the best visualization. Other times, you need a more complex visualization to illustrate your point. |
| Optimize the data-ink ratio | The data-ink entails focusing on the part of the visual that is essential to understanding the point of the chart. Try to minimize non-data ink like boxes around legends or shadows to optimize the data-ink ratio. |
| Use orientation effectively | Make sure the written components of the visual, like the labels on a bar chart, are easy to read. You can change the orientation of your visual to make it easier to read and understand. |
| color | There are a lot of important considerations when thinking about using color in your visuals. These include using color consciously and meaningfully, staying consistent throughout your visuals, being considerate of what colors mean to different people, and using inclusive color scales that make sense for everyone viewing them. |
| Numbers of things | Think about how many elements you include in any visual. If your visualization uses lines, try to plot five or fewer. If that isn’t possible, use color or hue to emphasize important lines. Also, when using visuals like pie charts, try to keep the number of segments to less than seven since too many elements can be distracting. |

## **Avoiding misleading or deceptive charts**

| **What to avoid** | **Why** |
| --- | --- |
| Cutting off the y-axis | Changing the scale on the y-axis can make the differences between different groups in your data seem more dramatic, even if the difference is actually quite small. |
| Misleading use of a dual y-axis | Using a dual y-axis without clearly labeling it in your data visualization can create extremely misleading charts. |
| Artificially limiting the scope of the data | If you only consider the part of the data that confirms your analysis, your visualizations will be misleading because they don’t take all of the data into account. |
| Problematic choices in how data is binned or grouped | It is important to make sure that the way you are grouping data isn’t misleading or misrepresenting your data and disguising important trends and insights. |
| Using part-to-whole visuals when the totals do not sum up appropriately | If you are using a part-to-whole visual like a pie chart to explain your data, the individual parts should add up to equal 100%. If they don’t, your data visualization will be misleading. |
| Hiding trends in cumulative charts | Creating a cumulative chart can disguise more insightful trends by making the scale of the visualization too large to track any changes over time. |
| Artificially smoothing trends | Adding smooth trend lines between points in a scatterplot can make it easier to read that plot, but replacing the points with just the line can actually make it appear that the point is more connected over time than it actually was. |

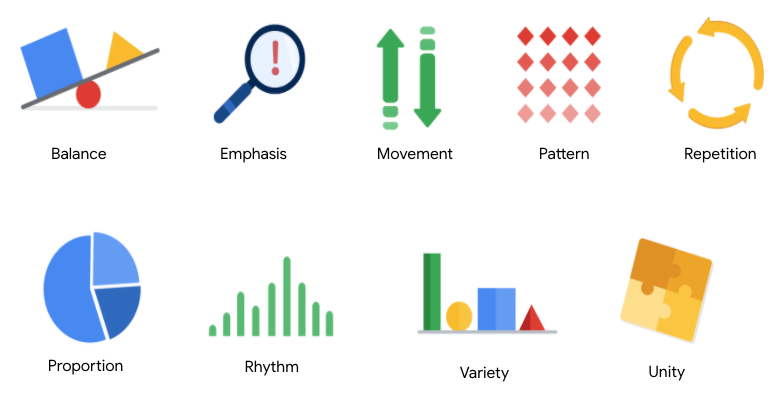
**How to choose charts for visualization:**

A decision tree is a decision-making tool that allows you, the data analyst, to make decisions based on key questions that you can ask yourself

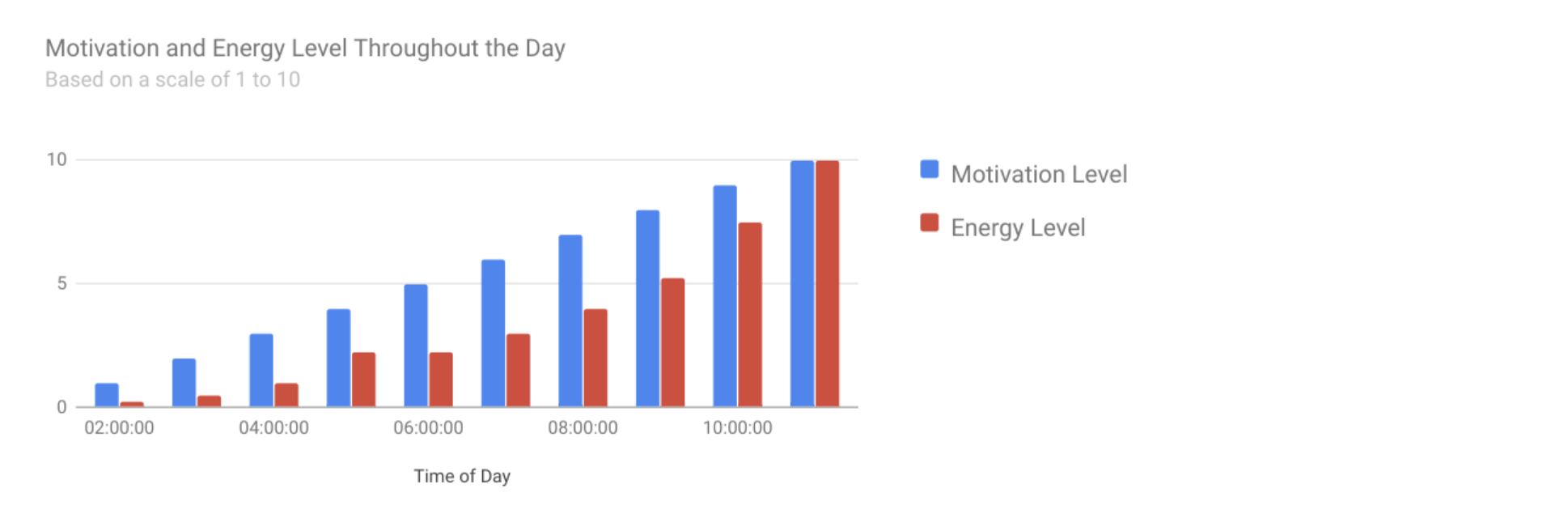


**Design Principles:**

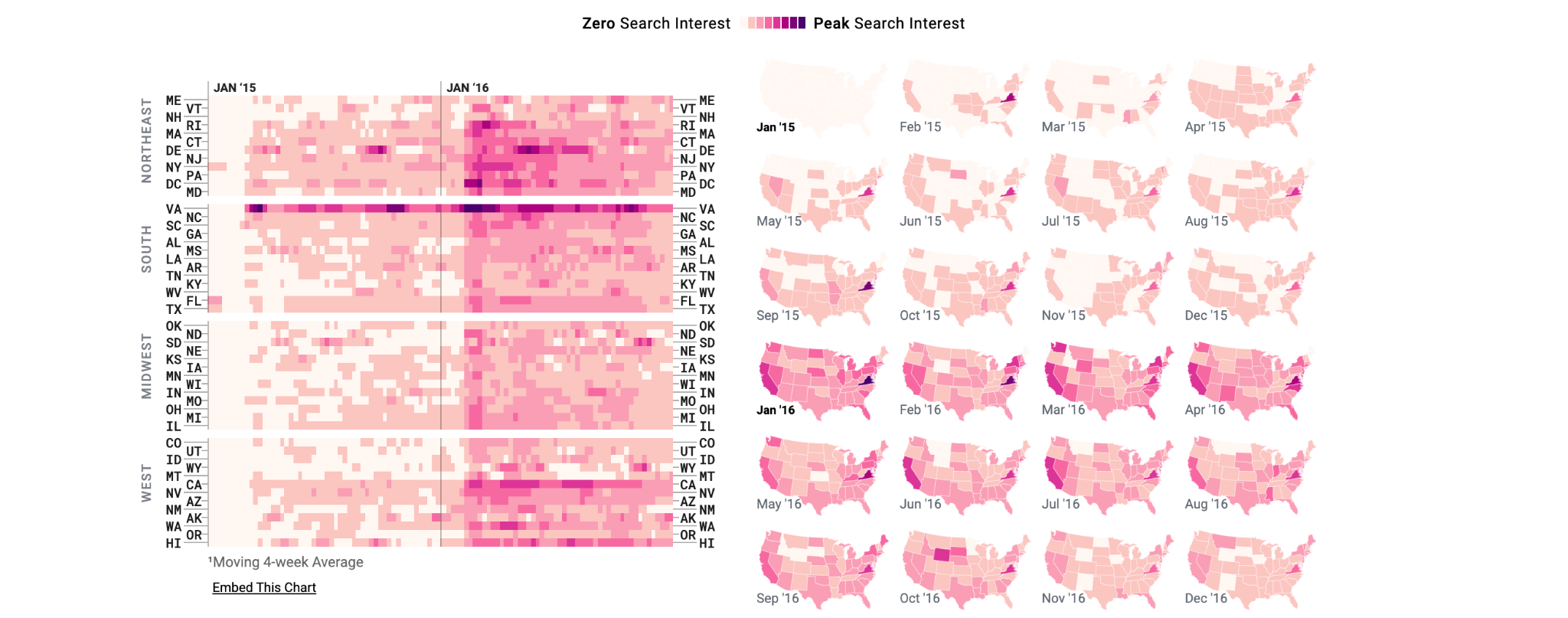
**Nine** principles of design are:



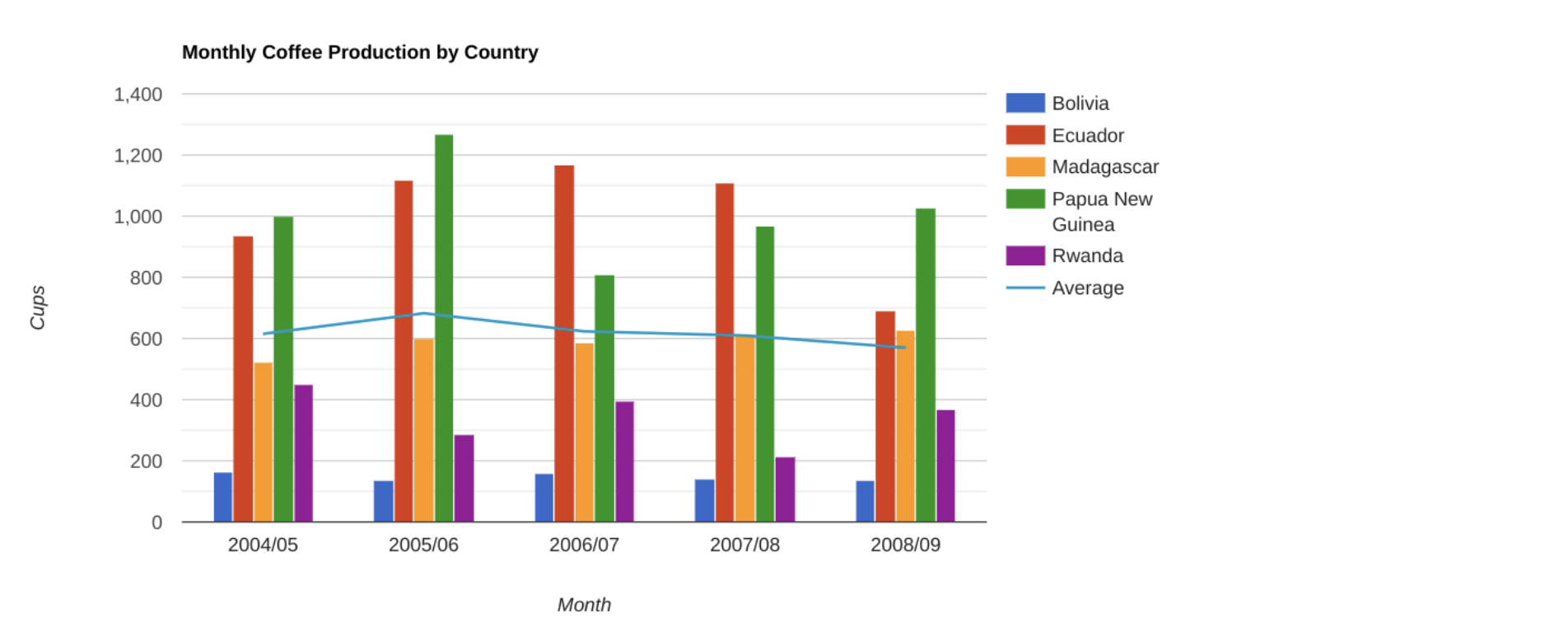
**1. Balance**: The design of a data visualization is balanced when the key visual elements, like color and shape, are distributed evenly



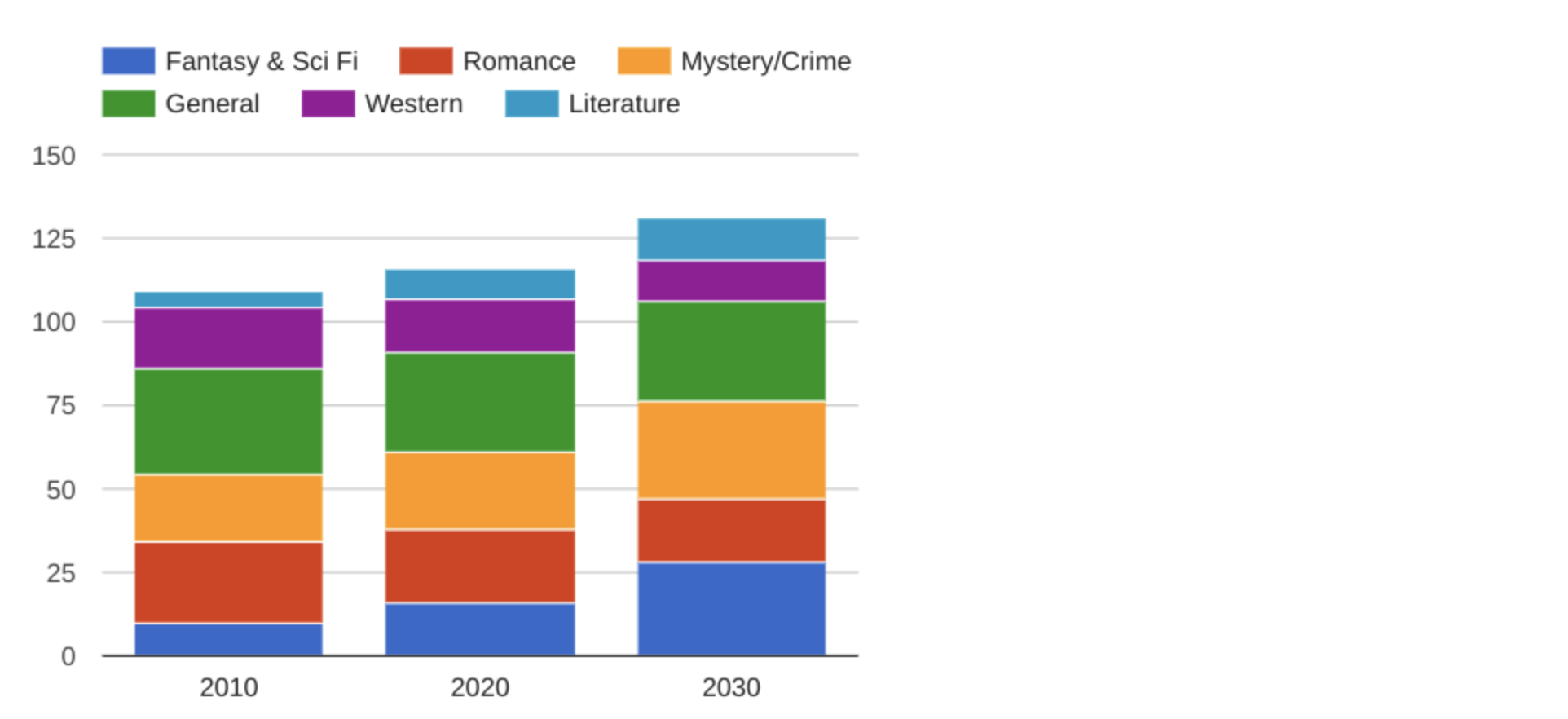
**2. Emphasis:** Your visualizations should emphasize the most important data so that users recognize it first. By using contrasting colors, you can make certain that graphic elements—and the data shown in those elements—stand out



**3. Movement:** Movement can refer to the path the viewer’s eye travels as they look at a data visualization, or literal movement created by animations.

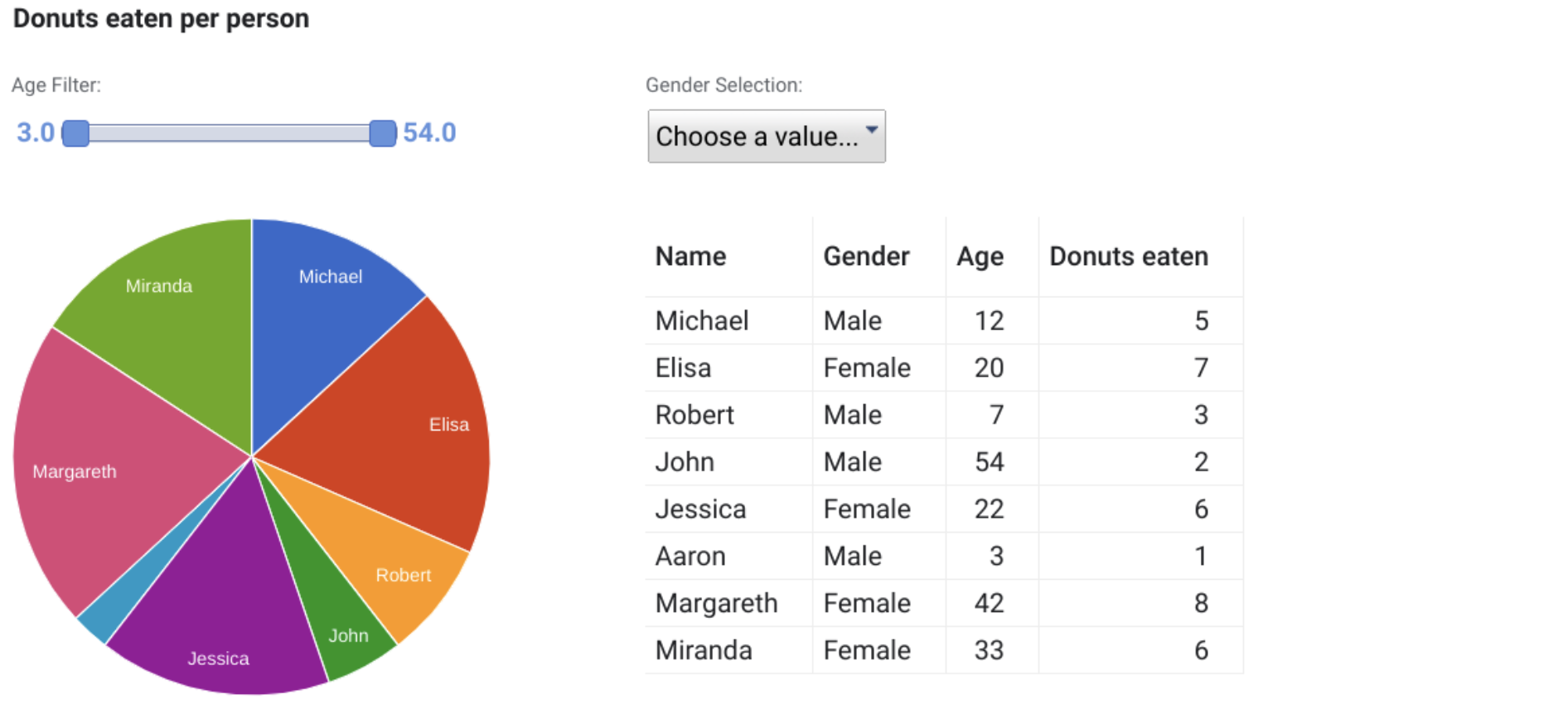


**4. Pattern:** You can use similar shapes and colors to create patterns in your data visualization



**5. Repetition:** Repeating chart types, shapes, or colors adds to the effectiveness of your visualization

**6. Proportion:** Proportion is another way that you can demonstrate the importance of certain data. It is important to make sure that each chart accurately reflects and visualizes the relationship among the values in it.



These first six principles of design are key considerations that you can make while you are creating your data visualization. These next three principles are useful checks once your data visualization is finished

**7. Rhythm:** This refers to creating a sense of movement or flow in your visualization.

**8. Variety:** Your visualizations should have some variety in the chart types, lines, shapes, colors, and values you use

**9. Unity:** The last principle is unity. This means that your final data visualization should be cohesive. If the visual is disjointed or not well organized, it will be confusing and overwhelming.

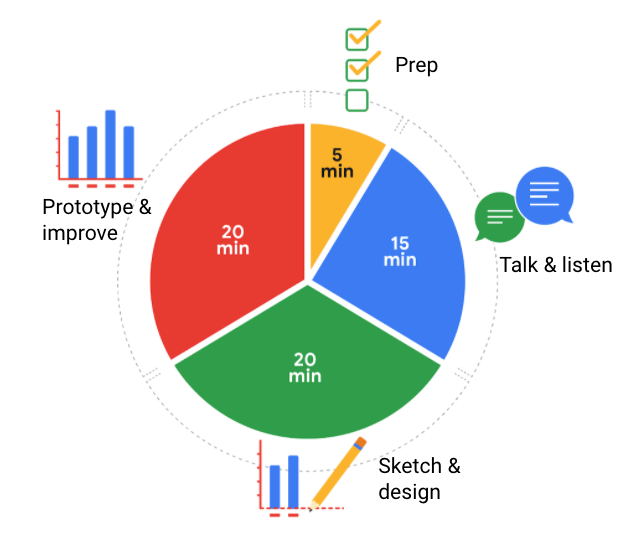
**Design Thinking:**

Design thinking for data visualization involves five phases:

1. **Empathize:** Thinking about the emotions and needs of the target audience for the data visualization
2. **Define:** Figuring out exactly what your audience needs from the data
3. **Ideate:** Generating ideas for data visualization
4. **Prototype:** Putting visualizations together for testing and feedback
5. **Test:** Showing prototype visualizations to people before stakeholders see them

| **Visualization**  **components** | **Guidelines** | **Style checks** |
| --- | --- | --- |
| Headlines | - **Content**: Briefly describe the data  - **Length**: Usually the width of the data frame  - **Position**: Above the data | - Use brief language  - Don’t use all caps  - Don’t use italic  - Don’t use acronyms  - Don't use abbreviations  - Don’t use humor or sarcasm |
| Subtitles | - **Content**: Clarify context for the data  - **Length**: Same as or shorter than headline  - **Position**: Directly below the headline | - Use smaller font size than headline  - Don’t use undefined words  - Don’t use all caps, bold, or italic  - Don’t use acronyms  - Don't use abbreviations |
| Labels | - **Content**: Replace the need for legends  - **Length**: Usually fewer than 30 characters  - **Position**: Next to data or below or beside axes | - Use a few words only  - Use thoughtful color-coding  - Use callouts to point to the data  - Don’t use all caps, bold, or italic |
| Annotations | - **Content**: Draw attention to certain data  - **Length**: Varies, limited by open space  - **Position**: Immediately next to data annotated | - Don’t use all caps, bold, or italic  - Don't use rotated text  - Don’t distract viewers from the data |

**Designing a chart in 60 minutes:**

****

**Data Storytelling:**

In data analytics, data storytelling is communicating the meaning of a dataset with visuals and a narrative that is customized for a particular audience

**Ex :** [Mapping New York's Noisiest Neighborhoods](https://www.newyorker.com/tech/annals-of-technology/mapping-new-york-noise-complaints)

**Live and Static Data:**

**Static data** involves providing screenshots or snapshots in presentations or building dashboards using snapshots of data. There are pros and cons to static data.

**PROS**

* Can tightly control a point-in-time narrative of the data and insight
* Allows for complex analysis to be explained in-depth to a larger audience

**CONS**

* Insight immediately begins to lose value and continues to do so the longer the data remains in a static state
* Snapshots can't keep up with the pace of data change

**Live data** means that you can build dashboards, reports, and views connected to automatically updated data.

**PROS**

* Dashboards can be built to be more dynamic and scalable
* Gives the most up-to-date data to the people who need it at the time when they need it
* Allows for up-to-date curated views into data with the ability to build a scalable “single source of truth” for various use cases
* Allows for immediate action to be taken on data that changes frequently
* Alleviates time/resources spent on processes for every analysis

**CONS**

* Can take engineering resources to keep pipelines live and scalable, which may be outside the scope of some companies' data resource allocation
* Without the ability to interpret data, you can lose control of the narrative, which can cause data chaos (i.e. teams coming to conflicting conclusions based on the same data)
* Can potentially cause a lack of trust if the data isn’t handled properly

## **Presentation**

## **Before the presentation**

1. Assemble and prepare your questions.
2. Discuss your presentation with your manager, other analysts, or other friendly contacts in your organization.
3. Ask a manager or other analysts what sort of questions were normally asked by your specific audience in the past.
4. Seek comments, feedback, and questions on the deck or the document of your analysis.
5. At least 24 hours ahead of the presentation, try and brainstorm tricky questions or unclear parts you may come across- this helps avoid surprises.
6. It never hurts to practice what you will be presenting, to account for any missing information or simply to calm your nerves.

## 

## **During the presentation**

1. Be prepared to respond to the things that you find and effectively and accurately explain your findings.
2. Address potential questions that may come up.
3. Avoid having a single question derail a presentation and propose following-up offline.
4. Put supplementary visualizations and content in the appendix to help answer questions.