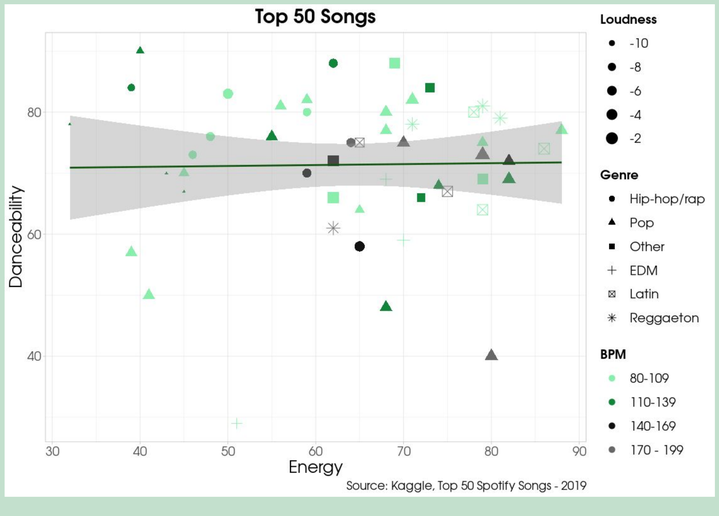
Statistical Graphics Principles: : **CHEAT SHEET**

Informative and Grouping/Search

An informative graph **clearly highlights the important trends** of pattern instead of attempting to display the relationship between all of the variables. The **title** or caption or body text **summarizes** these patterns in an **easy to understand** manner. This information can also be displayed using grouping and search, more specifically the **Gestalt Principles** of Design. These features are **easy to discriminate** between and allow variables to be mapped to separable dimensions for analysis.



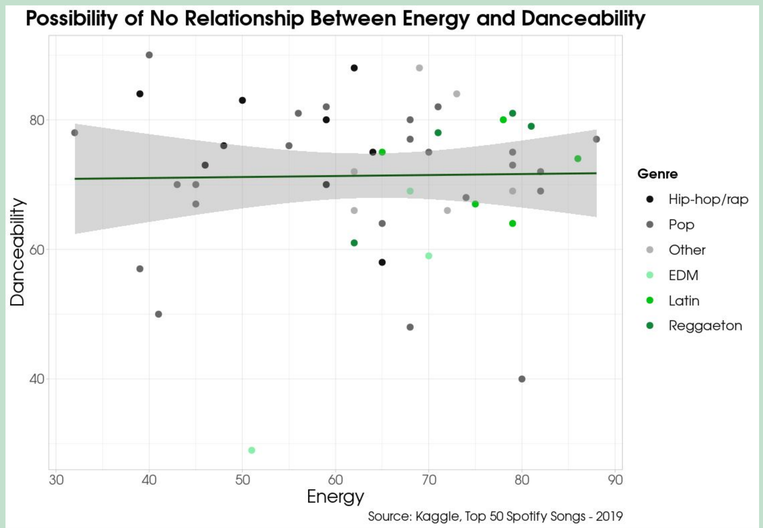
Source: Kaggle, Top 50 Spotify Songs - 2019

UNSUCCESSFUL PLOT

- An unsuccessful plot does not have an **informative title and caption** so it is unclear what the graph is showing
- The **pattern is not indicated** in the title, caption or text body
- It is difficult to find groups, follow lines making it **difficult to read**
- There are **too many aesthetics** being mapped to variables so distinct elements cannot be distinguished

SUCCESSFUL PLOT

- A successful plot has an **informative title** that summarizes a trend
- The variables plotted show an **interesting relationship**
- The **Gestalt principles** are **used effectively** to map variables to separable dimensions (this graph uses color)
- Elements are **aligned**
- Choice of **colors are easy to discriminate**



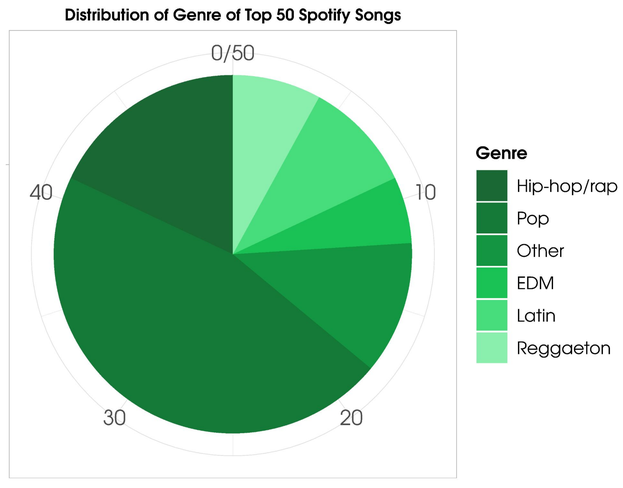
Source: Kaggle, Top 50 Spotify Songs - 2019

COPYRIGHT

Data Source: https://www.kaggle.com/leonardopena/top50spotify2019
Learn more from Alberto Cairo's Book The Functional Art

Cognition

In a successful plot, differences in quantities or proportions between categories of a variable should be **immediately noticeable** by the viewer. Also, categories should be **ranked by variables on which comparisons will be made**. An unsuccessful plot causes the user to compute differences mentally, which detracts from the mission of the graph. Additionally, the variables of interest in an unsuccessful plot are organized in a nonsensical order.



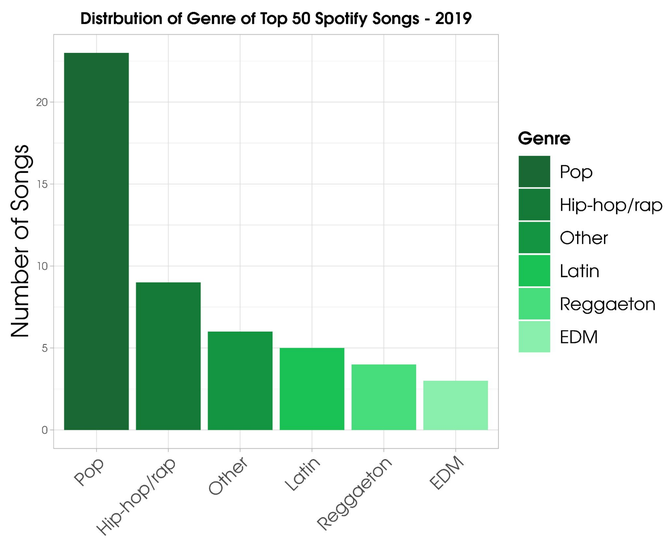
Source: Kaggle, Top 50 Spotify Songs - 2019

UNSUCCESSFUL PLOT

- According to the Gestalt Principles of Design, it is **harder for the human eye to comprehend differences in angle than almost any other measure** (position, length, etc)
- In this pie chart, the **utilization of angle** to represent the number of songs in each genre makes it **difficult to make comparisons** between genre groups, especially since some groups appear to be very close in size
- Additionally, the categories of genre in this plot are arranged in a **random order** such that those with similar angles are not placed near each other, **preventing side-by-side comparison**

SUCCESSFUL PLOT

- This bar chart uses **height** to represent the number of songs in each genre **makes it easier** to make comparisons between genre groups
- Arranging the categories in **decreasing order** is appealing to the eye and immediate comparisons between groups are able to be made



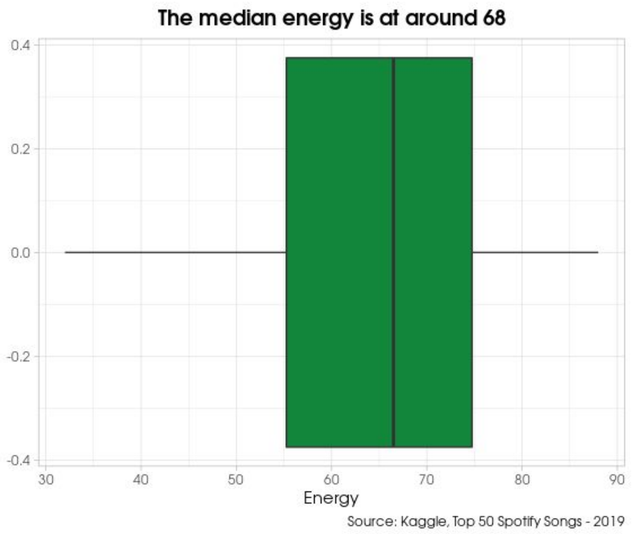
Source: Kaggle, Top 50 Spotify Songs - 2019

Statistical Graphics Principles: : **CHEAT SHEET**

Statistical Summaries

Statistical summaries supplement the underlying data by providing a big picture summary, showing overall trends, the global shape of distribution or any association between variables.

A successful plot **shows the underlying data**, with statistical summaries overlaid as necessary.

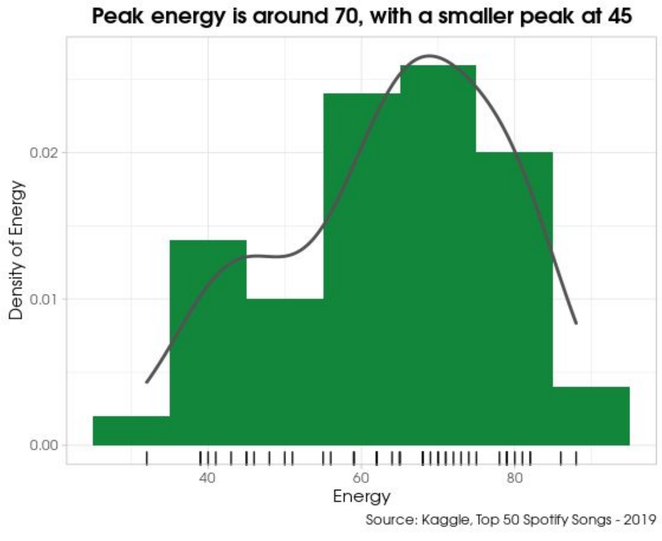


UNSUCCESSFUL PLOT

- An unsuccessful plot does not show the underlying data, and instead **provides just a summary of the shape**, like in the boxplot shown
- We may **miss local nuances** in the data like outliers or smaller peaks in the distribution as the statistical summary may be an **over-smoothing of the data**
- We can tell that there is a slight left skew in the distribution when looking at the boxplot. However, we miss the smaller peak at 45.

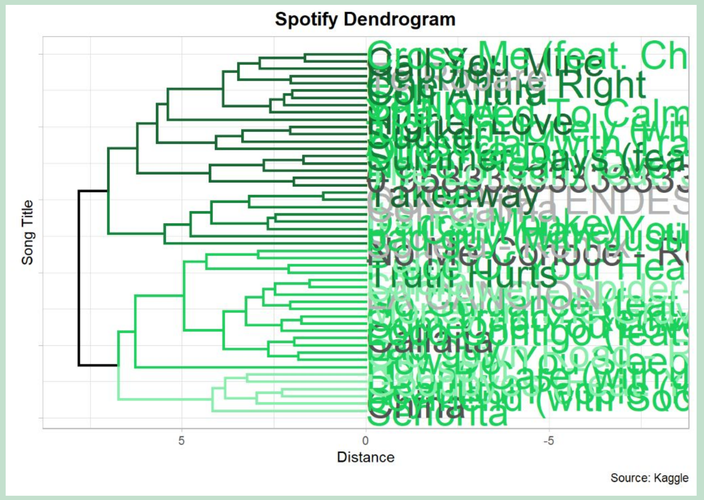
SUCCESSFUL PLOT

- A successful plot **overlays statistical summaries** over the original data
- Having the histogram and density plot reveals that there is **another peak at 45**, which we wouldn't find by just looking at the boxplot.
- Choosing an **appropriate bandwidth** for the density plot is also important to avoid over-smoothing and so we can see the **true shape** of the data.



Legible and Comprehensible

A successfully legible and comprehensible graph is **easily readable** (image format, color scheme, text size), with **informative** supplemental information (e.g. title, caption, axis labels, etc.)

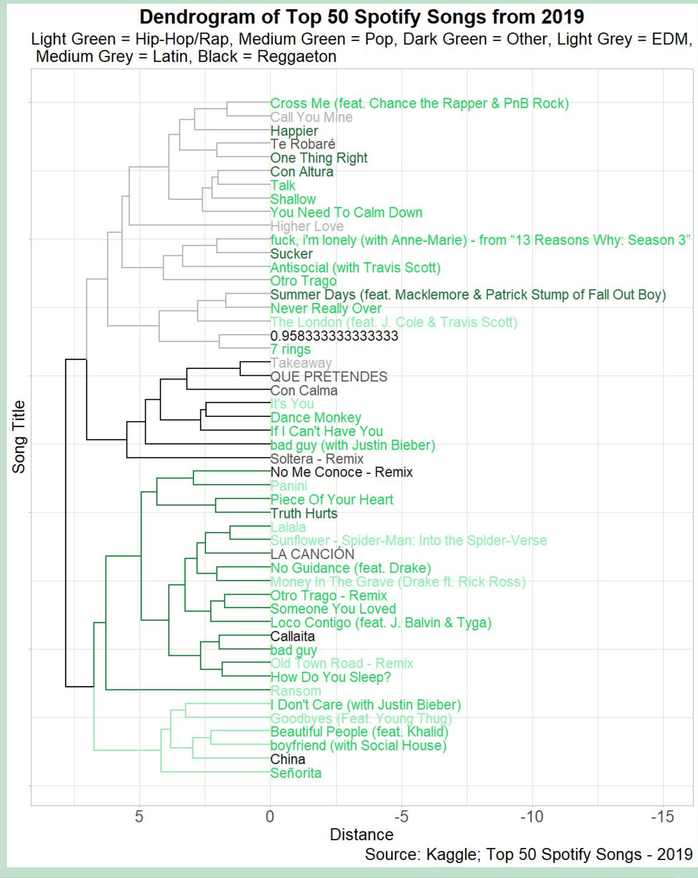


UNSUCCESSFUL PLOT

- An unsuccessful plot does not have an **informative title and caption** so it is unclear what the graph is showing
- It is **missing a legend** so it is unclear what the various colors of the track names mean, the colors are also **hard to distinguish**
- The **text not readable**: it overlaps with surrounding text, gets cut off on the edges of the graph, and the axes labels are too small
- The **image format** makes the graph blurry and difficult to read when stretched

SUCCESSFUL PLOT

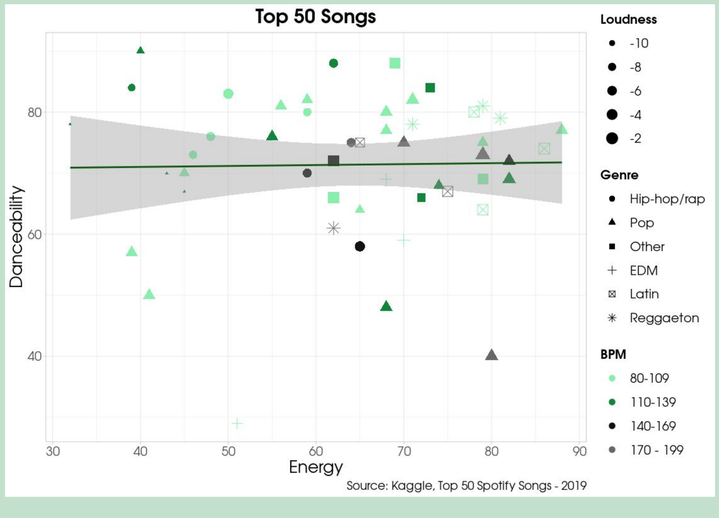
- A successful plot has an **informative title and caption** so it is clear what the graph is showing and where the data came from
- It is **has a legend** so it is clear what the colors of the track names mean
- The colors are easy to **distinguish** from each other
- The **text is a readable size** and does not overlap with surrounding text or get cut off on the edges of the graph
- The **image format** makes the graph clear and easy to read



Statistical Graphics Principles: : **CHEAT SHEET**

Informative and Grouping/Search

An informative graph **clearly highlights the important trends** of pattern instead of attempting to display the relationship between all of the variables. The **title** or caption or body text **summarizes** these patterns in an **easy to understand** manner. This information can also be displayed using grouping and search, more specifically the **Gestalt Principles** of Design. These features are **easy to discriminate** between and allow variables to be mapped to separable dimensions for analysis.



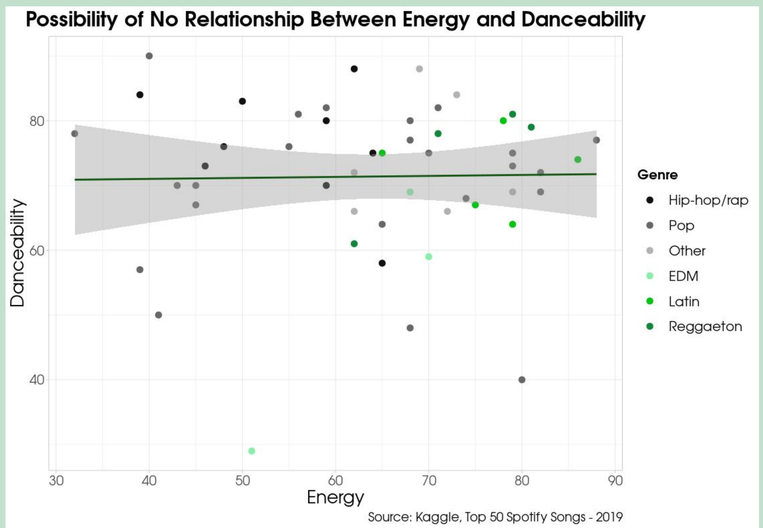
Source: Kaggle, Top 50 Spotify Songs - 2019

UNSUCCESSFUL PLOT

- An unsuccessful plot does not have an **informative title and caption** so it is unclear what the graph is showing
- The **pattern is not indicated** in the title, caption or text body
- It is difficult to find groups, follow lines making it **difficult to read**
- There are **too many aesthetics** being mapped to variables so distinct elements cannot be distinguished

SUCCESSFUL PLOT

- A successful plot has an **informative title** that summarizes a trend
- The variables plotted show an **interesting relationship**
- The **Gestalt principles** are **used effectively** to map variables to separable dimensions (this graph uses color)
- Elements are **aligned**
- Choice of **colors are easy to discriminate**



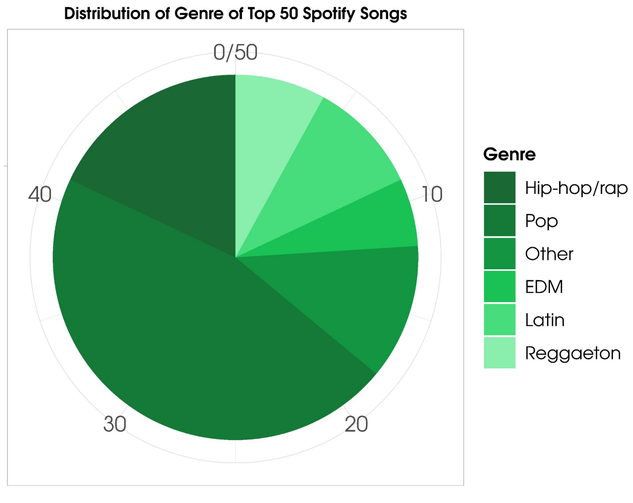
Source: Kaggle, Top 50 Spotify Songs - 2019

COPYRIGHT

Data Source: https://www.kaggle.com/leonardopena/top50spotify2019
Learn more from Alberto Cairo's Book The Functional Art

Cognition

In a successful plot, differences in quantities or proportions between categories of a variable should be **immediately noticeable** by the viewer. Also, categories should be **ranked by variables on which comparisons will be made**. An unsuccessful plot causes the user to compute differences mentally, which detracts from the mission of the graph. Additionally, the variables of interest in an unsuccessful plot are organized in a nonsensical order.



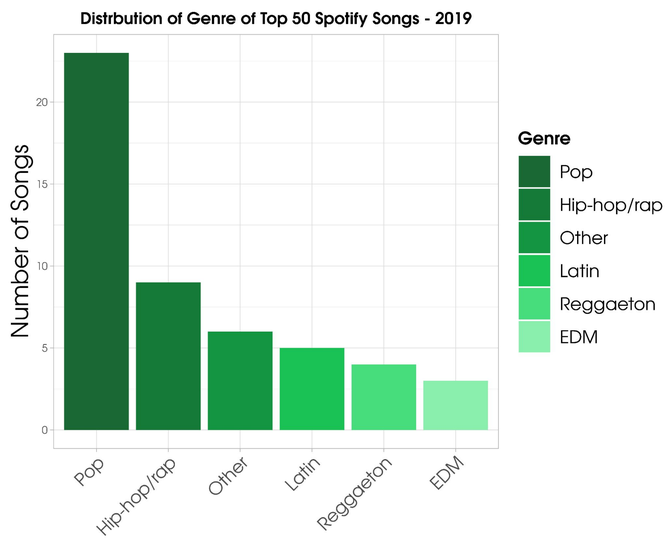
Source: Kaggle, Top 50 Spotify Songs - 2019

UNSUCCESSFUL PLOT

- According to the Gestalt Principles of Design, it is **harder for the human eye to comprehend differences in angle than almost any other measure** (position, length, etc)
- In this pie chart, the **utilization of angle** to represent the number of songs in each genre makes it **difficult to make comparisons** between genre groups, especially since some groups appear to be very close in size
- Additionally, the categories of genre in this plot are arranged in a **random order** such that those with similar angles are not placed near each other, **preventing side-by-side comparison**

SUCCESSFUL PLOT

- This bar chart uses **height** to represent the number of songs in each genre **makes it easier** to make comparisons between genre groups
- Arranging the categories in **decreasing order** is appealing to the eye and immediate comparisons between groups are able to be made



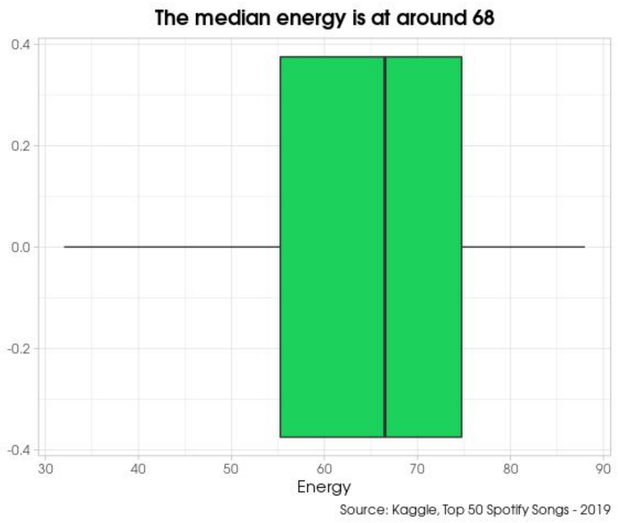
Source: Kaggle, Top 50 Spotify Songs - 2019

Statistical Graphics Principles: : **CHEAT SHEET**

Statistical Summaries

Statistical summaries supplement the underlying data by providing a big picture summary, showing overall trends, the global shape of distribution or any association between variables.

A successful plot **shows the underlying data**, with statistical summaries overlaid as necessary.

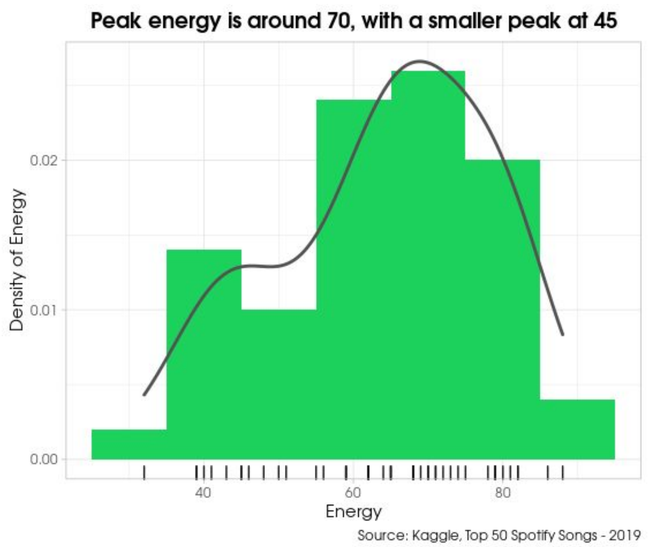


UNSUCCESSFUL PLOT

- An unsuccessful plot does not show the underlying data, and instead **provides just a summary of the shape**, like in the boxplot shown
- We may **miss local nuances** in the data like outliers or smaller peaks in the distribution as the statistical summary may be an **over-smoothing of the data**
- We can tell that there is a slight left skew in the distribution when looking at the boxplot. However, we miss the smaller peak at 45.

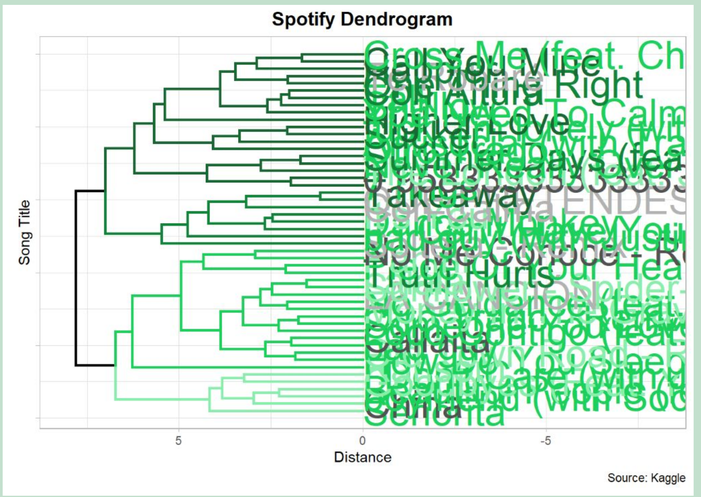
SUCCESSFUL PLOT

- A successful plot **overlays statistical summaries** over the original data
- Having the histogram and density plot reveals that there is **another peak at 45**, which we wouldn't find by just looking at the boxplot.
- Choosing an **appropriate bandwidth** for the density plot is also important to avoid over-smoothing and so we can see the **true shape** of the data.



Legible and Comprehensible

A successfully legible and comprehensible graph is **easily readable** (image format, color scheme, text size), with **informative** supplemental information (e.g. title, caption, axis labels, etc.)



UNSUCCESSFUL PLOT

- An unsuccessful plot does not have an **informative title and caption** so it is unclear what the graph is showing
- It is **missing a legend** so it is unclear what the various colors of the track names mean, the colors are also **hard to distinguish**
- The **text not readable**: it overlaps with surrounding text, gets cut off on the edges of the graph, and the axes labels are too small
- The **image format** makes the graph blurry and difficult to read when stretched

SUCCESSFUL PLOT

- A successful plot has an **informative title and caption** so it is clear what the graph is showing and where the data came from
- It is **has a legend** so it is clear what the colors of the track names mean
- The colors are easy to **distinguish** from each other
- The **text is a readable size** and does not overlap with surrounding text or get cut off on the edges of the graph
- The **image format** makes the graph clear and easy to read