Stat645: Week 3

Perceptual properties of continuous variables

Today, we're going to do some explorations into how we perceive different aesthetics, and look at a variation on the histogram that uses some of these ideas.

Perception

Start by loading `pop-selected.csv` in to R from the website. Last week we plotted it with putting time on the x-axis and rank on the y axis. Repeat that plot for reference. Instead of putting rank on the y-axis, we could display the track name and use another aesthetic to display the rank.

List possible aesthetics (other than position) that we could use to display rank, and then create plots using those aesthetics. What do you notice? What aesthetic makes it easiest to see the pattern over time? What makes it hardest? Does combining aesthetics make it better? What makes the original plot so good?

Use `scale_colour_gradient` to vary the colours at the high-end and low-end of the scales. Can you create a scale that makes it easier to see changes over time? Which end of the scale should draw your attention?

You've probably used size as one of the aesthetics - but what is size really mapped to? Create a small experiment to determine whether size is mapped to radius or area? Which do you think it should be mapped to?

Use scale_area to change the mapping so the square root of size is mapped to radius. How does this change the appearance of the plot?

Instead of points, we could use bars to display the data. `geom_bar` **has a default statistic that counts up observations, so if we want to use bars for this type of data we need to turn that off by specifying** `stat = "identity"`**:**

```
qplot(week, value, data = two, geom = "bar", stat = "identity") +
  facet_wrap(~ artist.inverted)
```

How does this display compare to the previous displays? What property are we reading for this plot?

What happens if you add on `coord_polar(theta = "y")`**? What property do we now perceive?**

Hanging rootogram

The hanging rootogram is plot developed by Tukey specifically for comparing a empirical distribution to a theoretical distribution, and is designed to answer questions like "does my data come from a normal distribution?" Read about the hanging rootogram on pages 312–315 of http://www.edwardtufte.com/tufte/tukey.

What are the important properties of the hanging rootogram and how do they make the desired comparison easier?

Why is overlaying a predicted density on top of a histogram not a good idea? Hint: http://www.michaelbach.de/ot/sze_sineIllusion/index.html

We are now going to construct our own hanging rootogram in R, using ggplot2.

Look at the code sample on the website. What do parts 1–4 do?

Use this data to create your own rootogram and hanging rootogram in R. For an extra challenge, try creating the suspended rootogram and add control limits. (Hint: `geom_hline`**)**

If you'd like to read more about the hanging rootogram, this article, http://www.jstor.org/stable/2683341, is a good start.