STAT7026 Final Project Part A

Graph 1: Joyplots!

- Creator: Henrik Lindberg
- Source: Twitter, https://twitter.com/hnrklndbrg/status/883675698300420098
- Codes: https://github.com/halhen/viz-pub/blob/master/sports-time-of-day/2_gen_chart.R
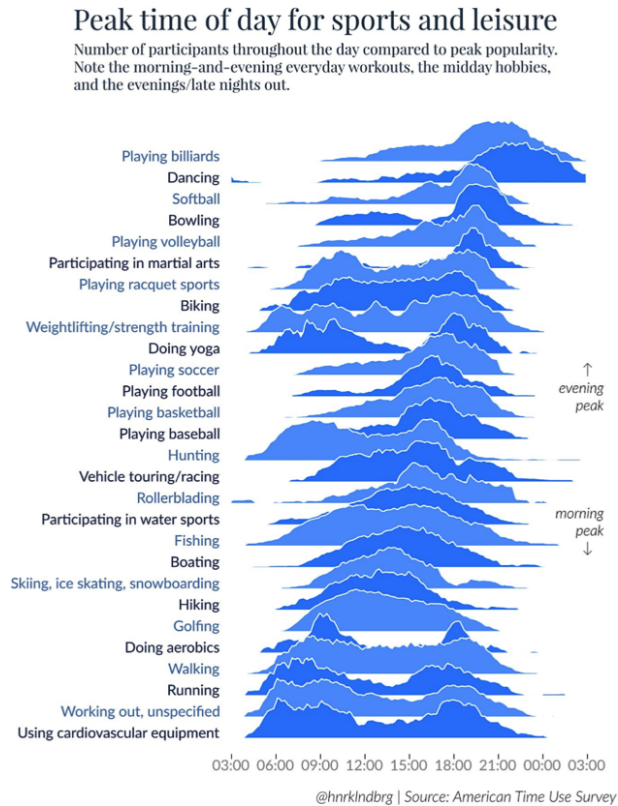
- Comments:
  - Although joyplot's predecessor aka *frequency trail* was first introduced by Brendan Gregg in 2013 [1], it was Henrik's tweet this year really made it popular again. The joyplot consists of a series of partially overlapping distribution plots, instead of traditionally plotting them in grids (inefficient for comparing), or overlapping in the same plot (pure madness in this case).
  - It sacrafices part of overlapped information in exchange of a clear comparison. The creator also uses two different tones of blue to strengthen the comparison. From a reader's view point, it feels like overlooking a range of mountains from a distance. *One cannot unsee what has been seen.* Intuitively, the reader would have an impression of which "peak" is the highest, etc.
  - Since parts of distributions are overlapped, one can never know a "valley" behind a "peak".
  - Inspired by Luis Carli's animated explainer [2], we believe a better version of joyplot is the one can modify viewing angle so that not only its advantage retains, but also the information loss is minimized.
 - Joyplots are now renamed as "Ridgeline Plots" and are available in the `ggridges` [3] package.

---

[1] Frequency Trails, http://www.brendangregg.com/frequencytrails.html.
[2] An Animated Gudie to Frequency Trails (aka Joyplots)
[3] ggridges: Ridgeline Plots in 'ggplot2', https://cran.r-project.org/package=ggridges.

Graph 2: A Day in the Life of Americans

- Creator: Nathan Yau
- Source: Flowing Data, A Day in the Life of Americans, http://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/
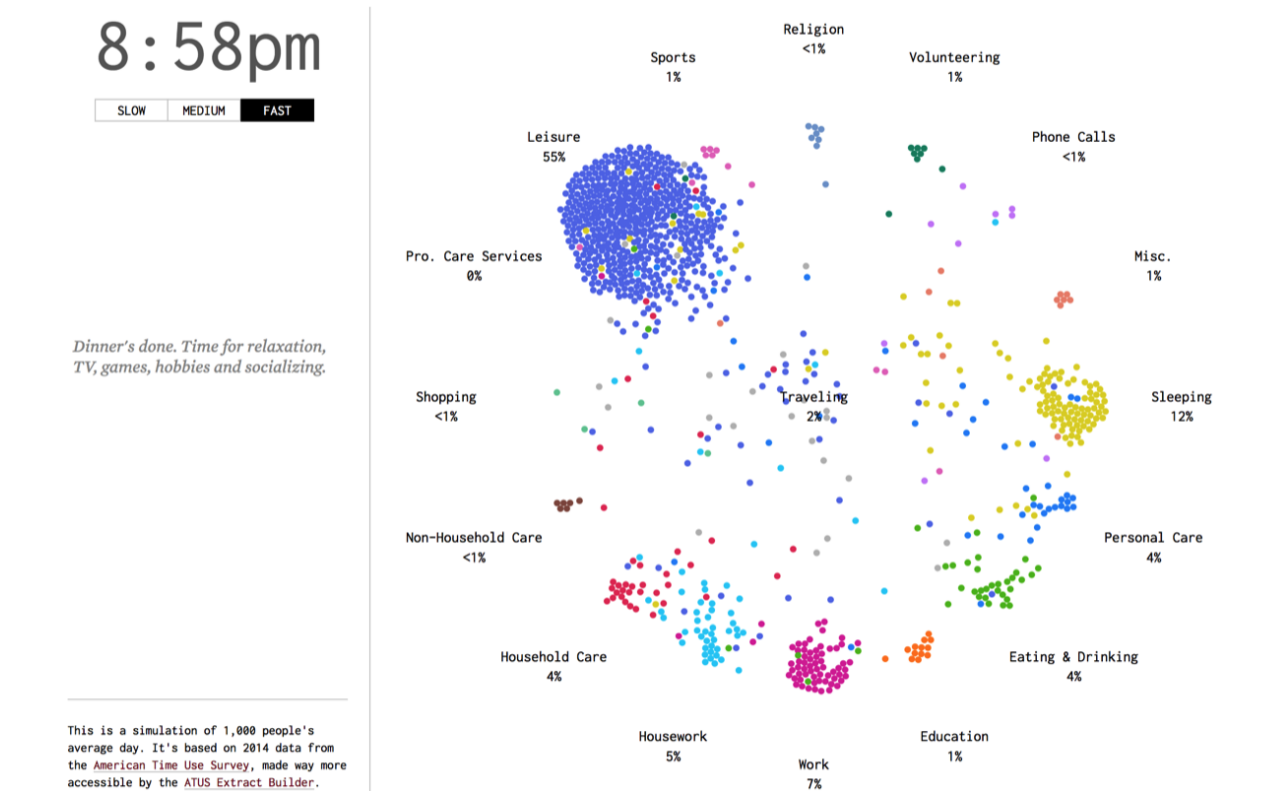
- Comments:
    - The highlight of this animated chart is that it visualizes time series data as moving dots. One major advantage of doing so is that, at a certain time point, we can directly know people are switching one activity to another.
    - Another benifit is, no matter what time it is, we can clearly determine what activities are "on trend".
    - Putting travelling in the center of this plot is a good choice since most of activities happen in different locations. A transition is temporary but necessary.
    - Also, an option on the left hand side controls the simulation speed.
    - An improvement could be accomplished if we use fewer types of activities. Some of them are really not that typically (e.g. volunteering), and some are included within others (e.g. making a phone call is usually job-related).

Graph 3: Violent crime in Houston

- Creator: David Kahle and Hadley Wickham
- Source: Kahle, David & Wickham, Hadley. (2013). ggmap: Spatial Visualization with ggplot2. R Journal. 5. 144-161.
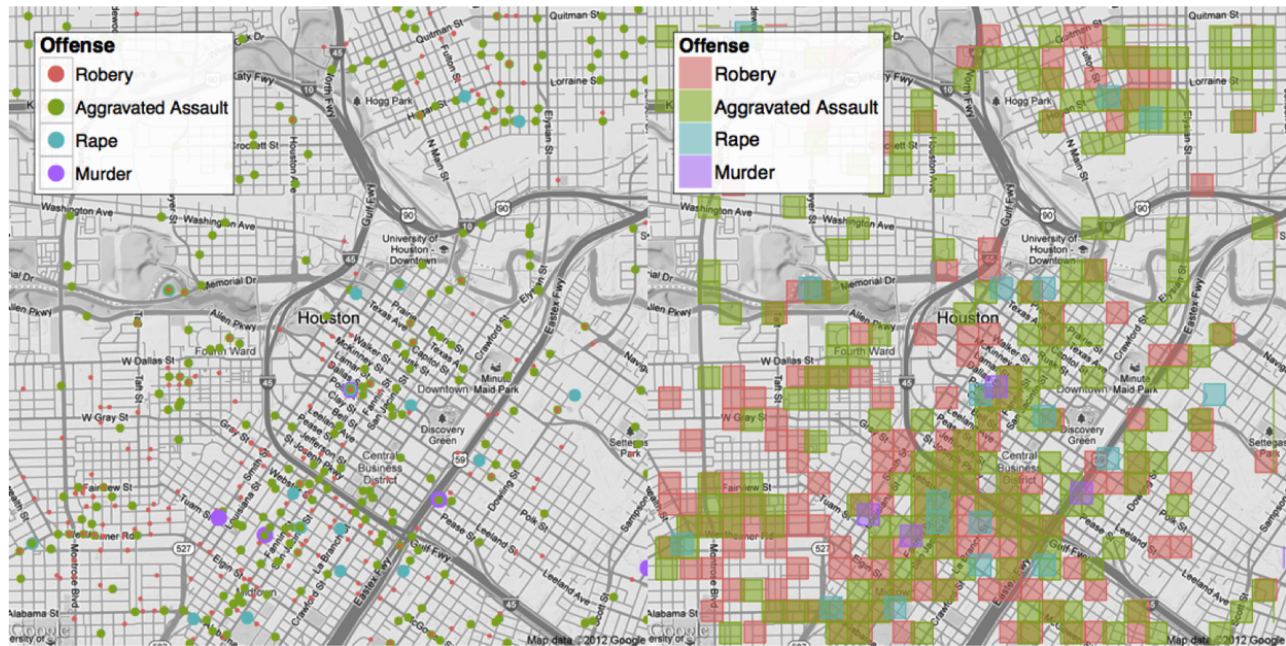
- Comments:
  - The best way to display spatial data is to plot them on a map.
  - Left:
    * Violent crimes are plotted in colored dots with different sizes. Larger points indicate the level of that crime is more severe.
    * It seems that Aggravated Assault is the most common violent crime in Houston, but this might be misleading. When the number of points plotted gets large, overlapping becomes a problem. Robbery in fact is a prevailing crime, but it is masked by Aggravated Assault's larger points.
  - Right:
    * When plotted as binned squares, the overlapping problem is alleviated. However, on the other hand, it lacks precision because a binned square is an aggregation of several crimes of the same type, so we won't know the exact location of a particular crime.
  - Basically spatial data plotting is a double-edged sword. If we want to precisely locate a data, we need to use points. If we really need to see the overall situation, a method of proximity is helpful.

Graph 4: Pop vs Soda

- Creator: Alan McConchie
- Source: The Pop vs Soda Page Project, http://www.popvssoda.com/
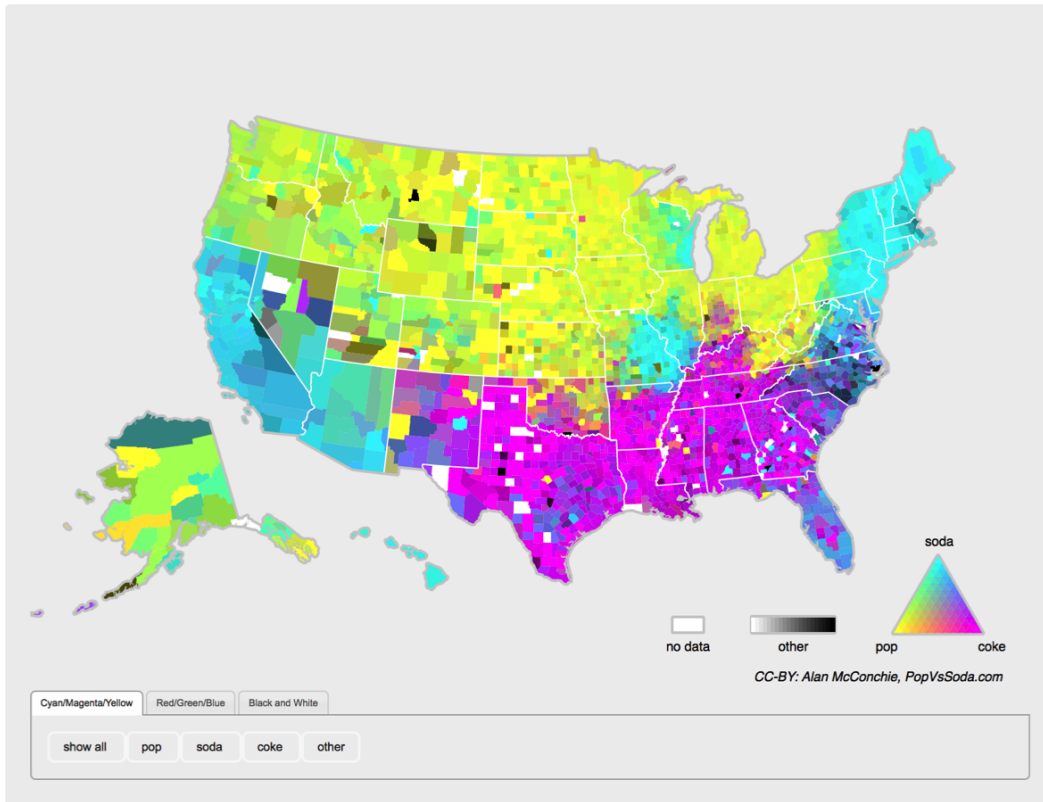
- Comments:
  - The creator uses three bright colors to represent three mainstream ways to call carbonated beverages.
  - The usage of gradient colors is definitely a pro, so the gradual transition from one major naming to another looks more consistent.
  - But the rich variety of colors in this map is also overwhelming. All those colors are bright colors. It makes looking at this map feel like watching a cheaply made fireworks in daylight. Also we find determining the naming in Alaska quite irritating since the colors of that part is not as extreme as others.
  - An improvement suggestion is actually embedded in the same page. If we switch the coloring from cyan/magenta/yellow to r/g/b, everything looks better.

Graph 5: Twenty Years of the NBA Redrafted

- Creator: Russell Goldenberg
- Source: The Pudding, Twenty Years of the NBA Redrafted, https://pudding.cool/2017/03/redraft/
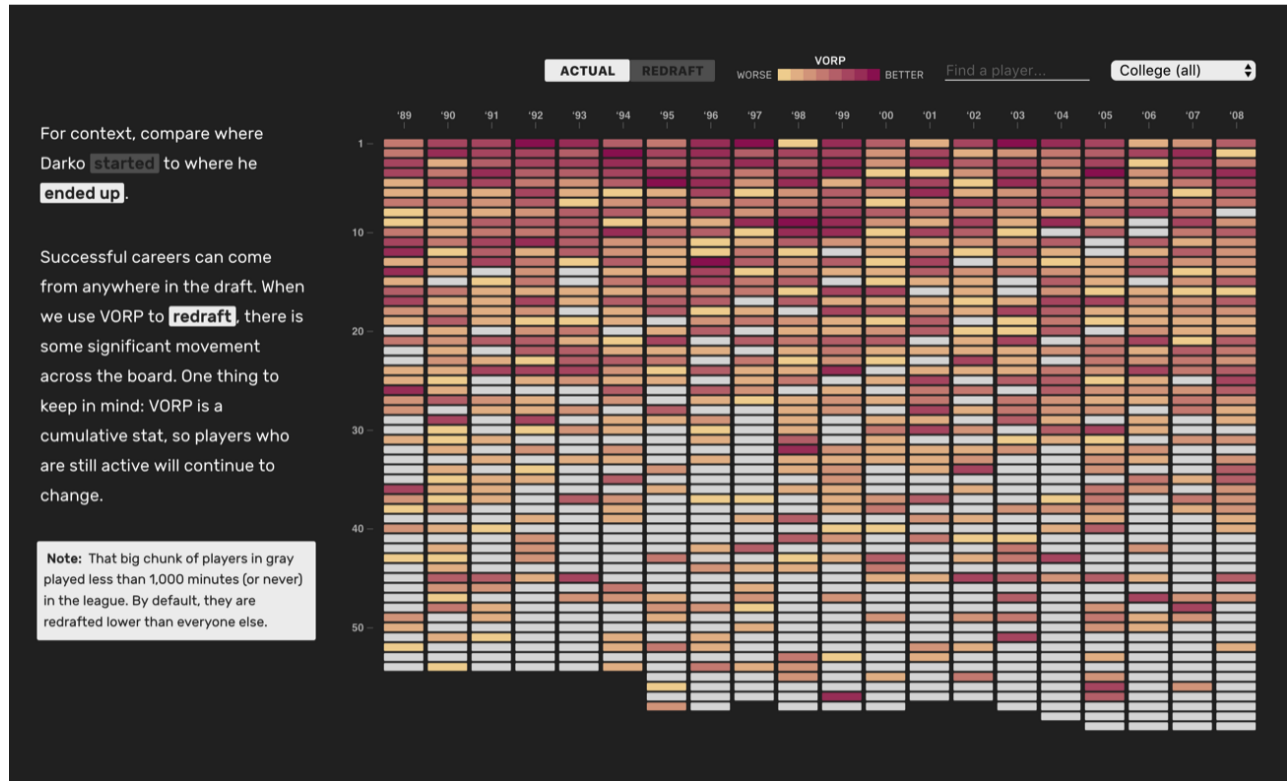
- Comments:
    - The heatmap of Value over Replacement Player (VORP) of NBA players underlines the discrepancy of reality and public's expectations of NBA players.
    - Generally, players drafted at a higher rank tend to have a more prosperous career (with relatively higher VORP). In fact, the redrafted version of this heatmap should be a perfect spectrum.
    - But the magic of this statistical graph is not telling you anything as a whole. It really lures you to click on a certain block and see the details of that player. One might wonder, why is this 2nd-round-pick player ended up with such a high VORP, or who is the bust in 1998's draft? Questions like this are the real goal of the data journalism, they simply encourage readers to raise questions and continue to read the article.
    - Some traditional graphs such as scatterplot and bar plots are also provided in the later part of the original article. And readers can answer their questions via reading the rest.