CS/INFO 3300; INFO 5100
Project 1
Final Report due 11:59pm Friday, March 26 (no slipdays)

In this project, you will create a **non-interactive data visualization as a team**. You will use d3, but the final result will be something that could be published on large-format paper or as a static PNG image on a web page. **Your final visualization must be dynamically generated by d3 when the page is loaded.** We will grade your work as it appears when we load the HTML file in a browser. **We will not click, hover, or interact with anything. We will not view animations or transitions. At most, we will scroll down a page** (which can create a kind of interaction in itself). You are encouraged but *not* required to integrate multiple data sources for this project. Finally, you will write a **final report** on your design and implementation. You will turn in your code via CMS along with your final report. Groups are encouraged to create a **Cornell GitHub repository** for their project to make collaborating on code easier and add accountability.

This project is intentionally very open-ended. It is up to your team what data you will visualize and how you will go about using D3 to realize your design. When thinking about your project, you ought to think in terms of the **insights you want to convey through your visualization** rather than displaying the most points or the most attributes possible. What points or trends do you want viewers to notice or what argument do you want to make? How will viewers *navigate* your visualization (visualization has compositional lines just as art does)? Successful projects tend to focus on one core message and align all design elements (marks, visual channels, annotations) towards helping convey that message.

All groups are encouraged throughout the project to seek feedback from fellow classmates and course staff. The best way to improve a visualization is to have others critique it. You will be required to submit **2 milestone responses** to CMS which will be graded on a completion credit basis. These milestones will be a chance to get some feedback on your progress.

Examples and sources: You are encouraged to find data and inspiration from other sites. Make sure you acknowledge these in comments and in your written description. If you choose to scrape data, please scrape respectfully and ethically. Any code that you did not write yourself (such as d3) must go in a separate .js file. **Unacknowledged code or concept reuse will be handled with standard academic integrity procedures.**

Regarding grading: This is an **open-ended assignment**. With homework we have a specific idea of what we want and we "take off" points when your work deviates from that. The reason project experience is the single most valuable asset you can bring to a new job is that we do *not* have specific ideas about what projects should look like: **it's up to you and your teammates**. As a result, think of the criteria below as an opportunity to "earn" points, not "lose" points. Our

principle with projects is that better work should get better grades. That does not mean that we curve: there's no reason we wouldn't in theory give everyone 100s, but in practice "perfect" grades are extremely rare.

Regarding teamwork and conflicts: If you have concerns about how your team is working contact a graduate TA individually as soon as possible. In rare circumstances we will differentiate grades with a group, but we are much happier to help a group succeed.

Turning your project in

You will upload a final zip file containing your project code and final report to CMS. Graders will then run a web server from the project root directory in order to view your submission. **Tit is absolutely critical for you to turn in correct, working files. Please re-download and test your project file prior to the deadline to verify you have submitted what you intended. We suggest that you elect one group member to turn in the file via CMS and have the other group member(s) download and check the contents.**

Milestones

Submit a written status report, one per group, to CMS by 11:59pm on the following dates. On-time completion of these reports will be worth **6 points in total**.

- **Friday, March 12**:
  - List five ideas you have for your visualization project (1-2 sentences each).
  - For each member, list their assigned tasks for the following week.
  - If you have any questions / requests for feedback, feel free to include them and provide sketches or drawings.
- **Friday, March 19**:
  - Begin by summarizing your final project idea in 1-2 sentences.
  - For each team member, write 1-2 sentences describing how they contributed to the project during the week.
  - For each member, list assigned tasks for the next week prior to the deadline.
  - If you have any questions / requests for feedback, feel free to include them and provide sketches or drawings

(next page)

Grading criteria

Your final submission has two parts, a **d3-based static data visualization (55 pts),** a **written description of your visualization (35 pts)** which includes an **outline of team-member contributions to the project (4 pts)**. (The final 6 points come from meeting milestones)

1. **The root directory of your project must include** a**n HTML page called index.html containing your visualization.** Include any additional script files and any additional data files in your repository, preferably in JSON form. Make absolutely certain that you have included all data files necessary and that your paths work properly if a web server is run from your project root directory. We will not be able to grade projects that do not run. You may import the d3 library from d3js.org, but all other libraries and data files must be stored in the repository. Your visualization will be graded on the following elements:

    a) **Complexity of the data**. Find a dataset that is manageable, however you ought to avoid trivial data. There should be more than two variables, for example. An advanced project might combine multiple datasets to provide a unique, novel perspective. Editing is important! Beginning projects often have too little data or too much. Don't overwhelm us with information.

    b) **Technical correctness**. The code must actually do what you intend it to do. We also prefer good style in coding: use informative variable names, consistent indenting and whitespace, and informative comments.

    c) **Creativity**. Beginning projects often look like online examples or things we've seen before. Advanced projects will make us think "how did they do that?" or use something familiar in an unfamiliar way. Don't be boring.

    d) **Proper use of visual channels**. Use scales such as position, shape, color, and text appropriately for variables. Advanced projects give us accurate impressions of the underlying data values, allow us to make comparisons between relevant data points, and balance between focus and context. Beginning projects are often hard to interpret and make comparisons difficult.

    e) **Usability**. Someone viewing your work should be able to understand the data values represented in the visualization easily and accurately. Advanced projects make choices that are clear and intuitive and may walk us through specific examples. Beginning projects often leave us wondering what we're looking at or make us read long descriptive paragraphs to figure out what's going on.

    f) **Overall polish**. Beginning projects will look like a collection of parts, with default styles. Advanced projects will have a sense of unity, even if they have multiple sections.

    g) **Motivation**. What's the point? What are you trying to say? Beginning projects will present information. More advanced projects will have a clear argument and use carefully chosen combinations of marks and channels to guide our attention to the evidence that supports that argument. Advanced projects deliver insights.

2. **Also include in the root directory of your project a PDF file containing a written description of your project**. There are no specific page / word limits, but 2-4 pages usually are sufficient. This document should contain:

    a) **At least one screenshot** of your final, static visualization.

    b) **A description of the data**. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. **Editing is important! You are not required to use every part of the dataset.** Selectively choosing a subset can improve usability. Describe any criteria you used for data selection.

    c) **An overview of your design rationale.** A good rule of thumb to follow is "**every pixel must be justified."** Instead of a 100,000-element breakdown, give us an overview of the design decisions you made and the trade-offs inherent in how you displayed the data. **This part ought to include a description of the mapping from data to visual elements**. Describe marks and channels you employ such as position, color, or shape. Mention any transformations you performed, such as log scales.

    d) **The story**. What does your visualization tell us? What was surprising about it? What insights do you want to convey to the viewer of your visualization?

3. At the end of your PDF file, also include an **outline of team contributions to the project**. Identify how work was broken down in the group and **explain each group member's contributions to the project**. Give a rough breakdown of how much time you spent developing and which parts of the project took the most time.

Best practices

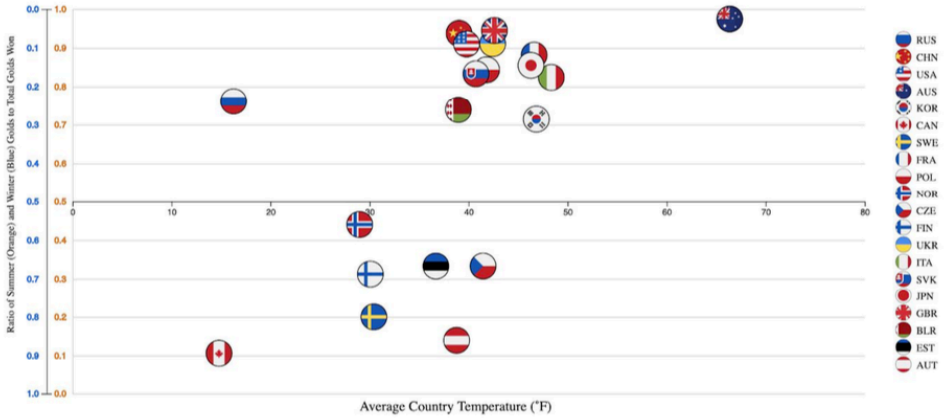
- Start now.
- Talk to each other. Listen and value each other's different perspectives.
- If a group member will be unavailable for any period during the project, figure out in advance how you will work around that absence.
- While we have attempted to group students based on shared time zones, it will not be a perfect matching. Discuss how you plan to collaborate remotely on the project early.
- Use relative paths for data, images, and other resources: do not start URLs with "/". Your project will be one directory among many, not the document root.
- Set up a code repository, like Github or Bitbucket. This is always a good idea, but it can also provide insurance if something goes wrong with your CMS submission.
- No, seriously, start now.

Example projects:

Visualizing Olympic medal winning stats with respect to country temperature

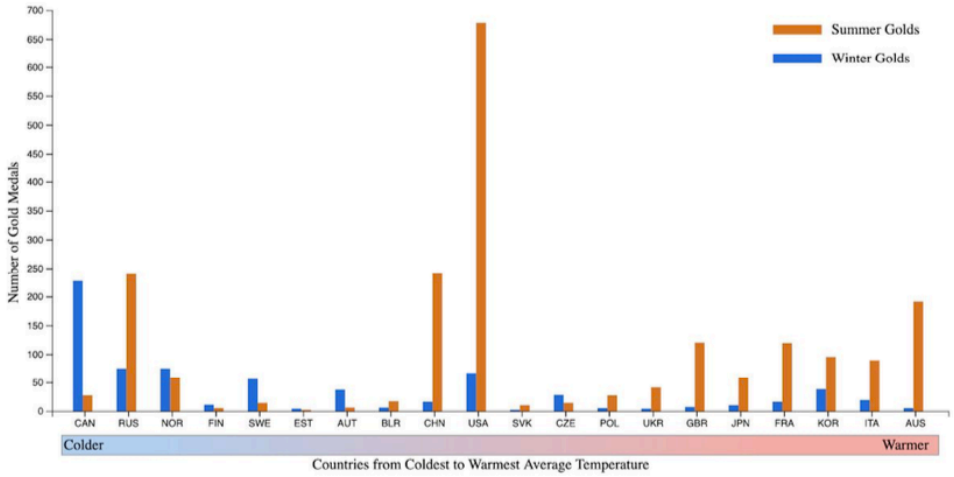
Graph 1: Visualizing the Olympic gold medal wins for countries of varying tempteratures

We want to see if the average temperatures of countries play a role in whether or not they are better at summer Olympics or winter Olympics. Below is a visualization of countries based on their average temperature and their proprtions of summer and winter golds to their total golds. We want to see if colder countries have a higher proportion of winter golds and warmer countries have a higher proportion of summer golds.



Graph 2: Visualizing the number of summer and winter Olympic golds for each country

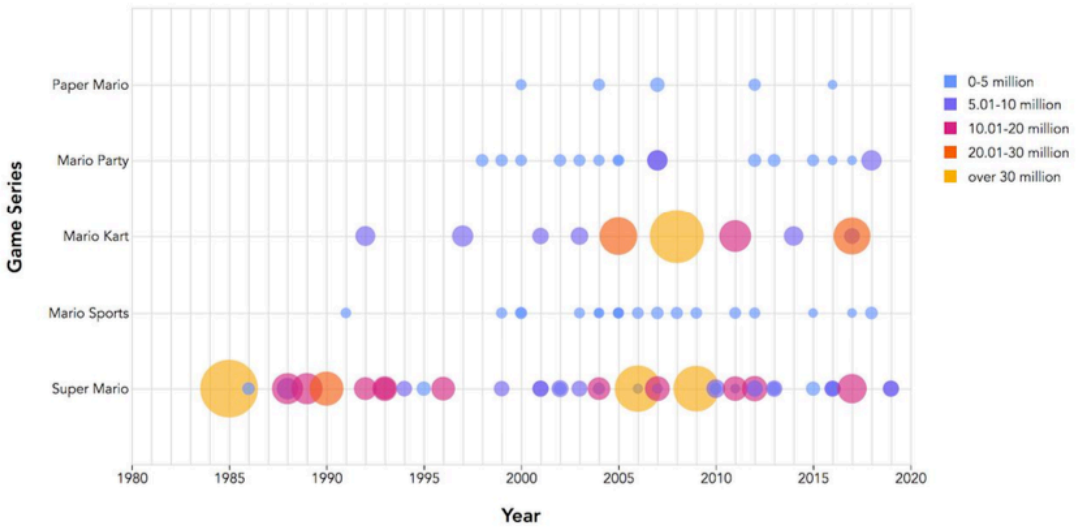
This visualization offers a numerical analysis of the number of summer and winter golds each country won from 1996-2014. While the first scatterplot primarily dealt in ratio of a certain gold to total golds, this graph hopes to convey the ratios through numbers, to compare the relative number of total medals between countries.



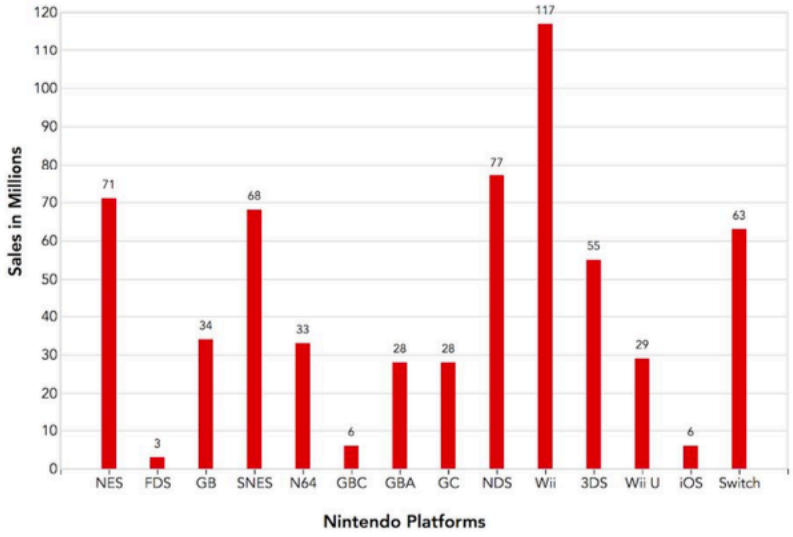
Investigating sales of Mario franchise video games

(for the record, Super Mario World is the best of them with Super Mario Kart 64 in close second -JRz)

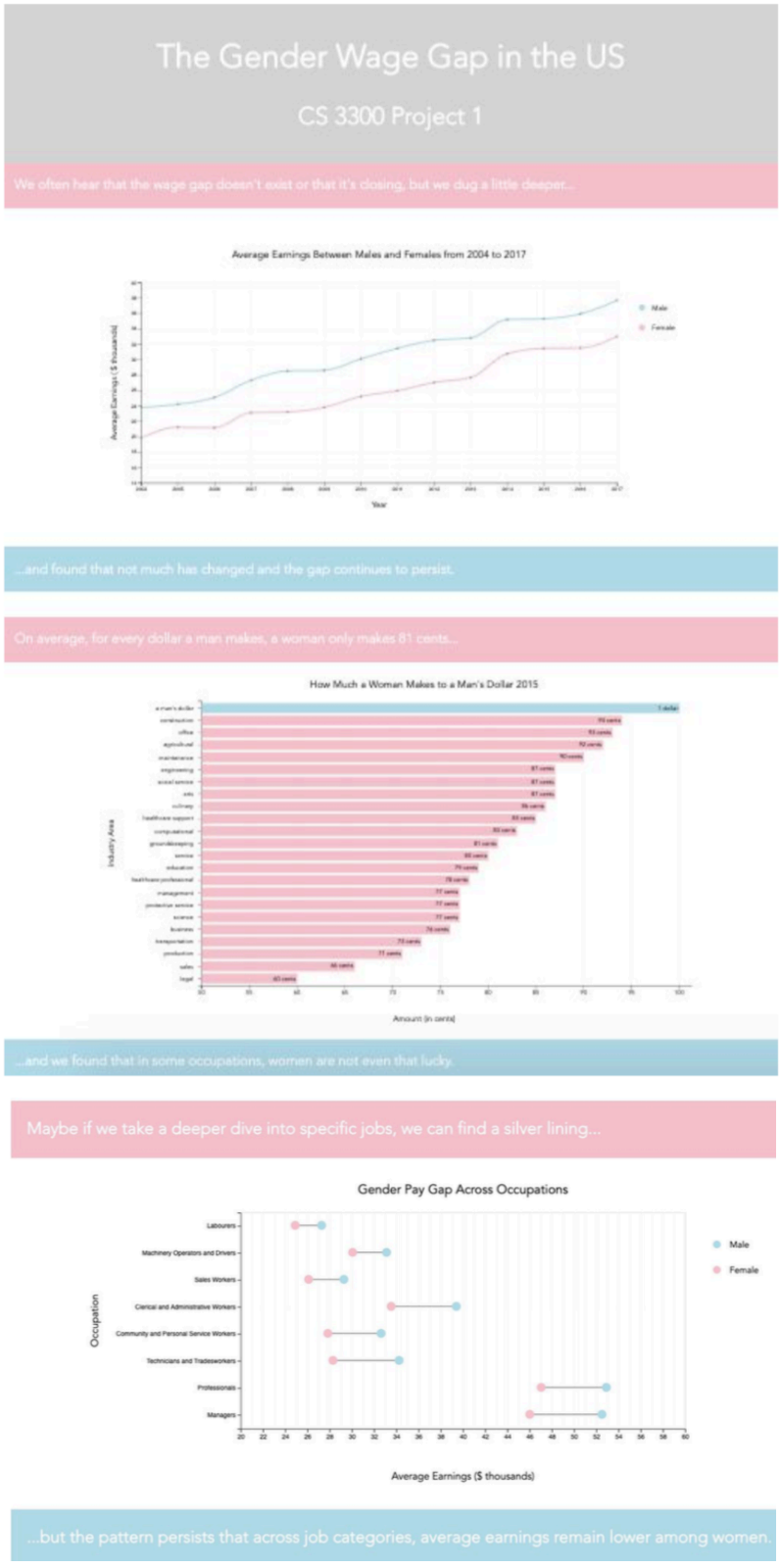
Top Selling Series Over Time



Game Sales by Console



Exploring the gender wage gap in the United States
(notice the use of scrolling to tell a story about the data)



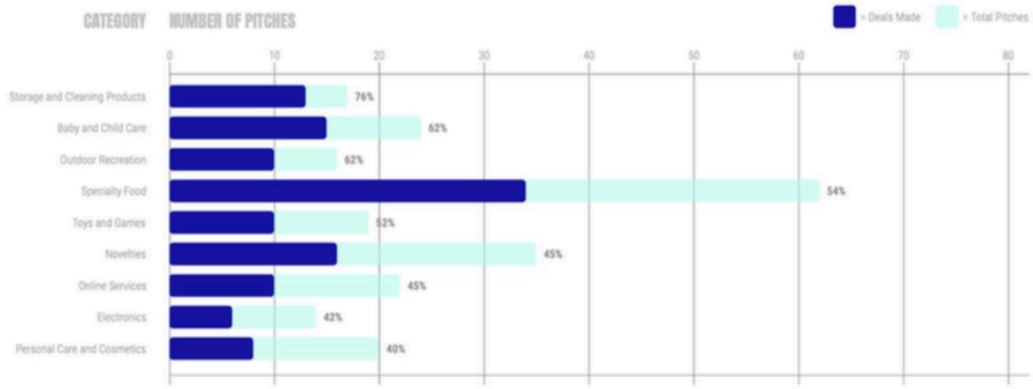
An analysis of startup pitches on the TV show Shark Tank

Breaking down idea **categories** across seasons one through six.

Since 2009, Shark Tank has played a major role for budding entrepreneurs and early startups to publicize their ideas. Each team holds a pitch in front of "Sharks" to convey their products, in hopes of winning an investment deal. Breaking them down by category, we observe the differences among company valuation, equity offer, amount asked for, and if a deal was made.

DEALS VS. REJECTIONS

The **Top 9** most popular categories for Shark Tank pitches had varying success rates, with an overall average of **53%** of pitches getting a deal. The most common category is **Specialty Food** (62 pitches), with a **54%** success rate, while **Storage and Cleaning Products** (17 pitches) had the highest success rate out of these top nine, at **76%**.



VALUATION VS. EXCHANGE FOR STAKE

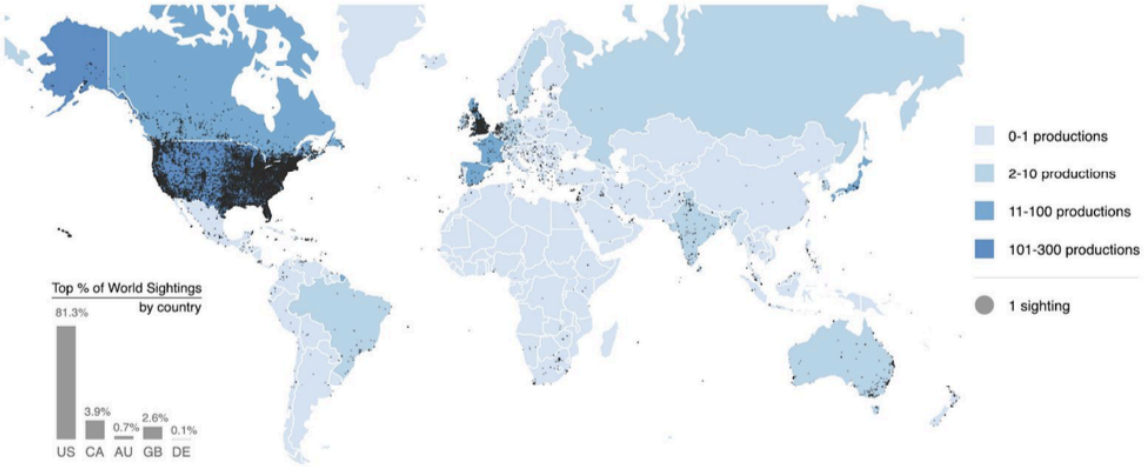
Mobile Apps had the highest average valuation of **$10,700,000** with an average exchange for stake of **6.25%**. Golf Products had the highest average exchange for stake of **31.5%** with an average valuation of **$390,380**. Specialty Food had the highest number of entries, topping at **62**, with an average valuation of **$1,791,717.41** and exchange for stake of **17.26%**.

SORTED BY HIGHEST AVERAGE VALUATION

Comparing Sci-fi movie releases with UFO sightings

Science Fiction Movie Productions Per Country and UFO Sightings Around the World, 2000-2013

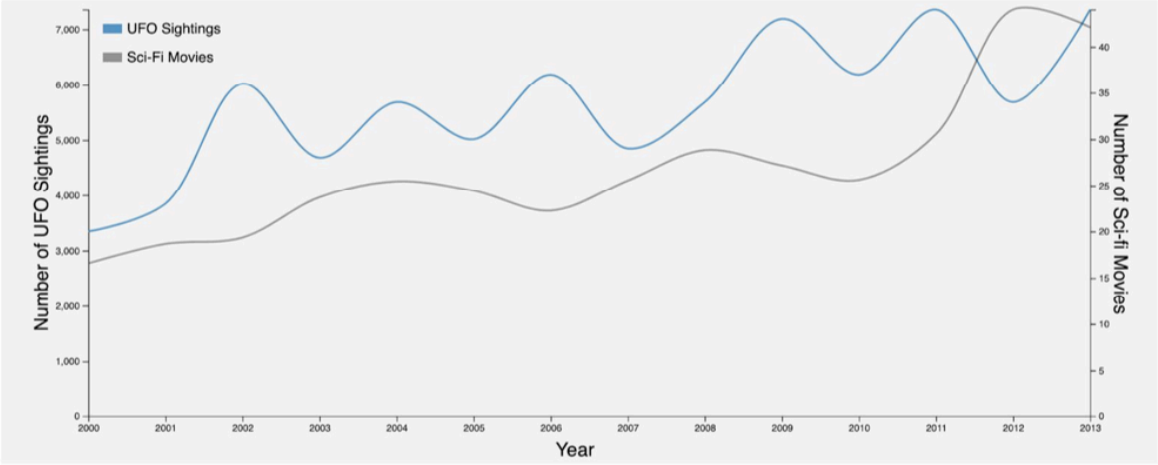
This map shows the total movie productions per country that have been categorized as science fiction and the locations of all recorded UFO sightings for a thirteen year time period.



The countries that have a higher rate of science fiction movie production also have a higher percentage of the world's UFO sightings. 60.7% of the world's science fiction movies are produced in the United States and 81.3% of the world's UFO sightings also occur within the United States.

Growth Trend of UFO Sightings and Sci-fi Movies, 2000-2013

We are making more and more sci-fi movies in the new century, and we love them. But, hmm, are we also seeing more and more UFOs? This timeline shows a general positive relationship between sci-fi movies released and UFO sightings over a time period of thirteen years.



Analyzing "random acts of pizza" sub-reddit user behavior