Data Visualization

Fall 2017

Matthew Turk

Usernames

goo.gl/PXjUVQ

Basics

1PM-3:50PM Fridays, LIS-126

Matthew Turk - mjturk@illinois.edu

Office Hours: Wednesday 2-4, LIS 222

TA Jarai Carter - carter31@illinois.edu

http://github.com/UIUC-iSchool-DataViz/

Timed Activity

Each of you has a notecard with a number.

On the back of that notecard:

- What are the most memorable movies you saw over the last year?
- Do you prefer cats or dogs?
- How would you quantify your experience in visualization?
- What would you guess the square footage of this building is?
- People per row in this class.

Break into groups based on your numbers, and visualize the results by whatever method you choose: by hand, by computer, or otherwise.

Affix to or inscribe upon your sheet of paper.

Row 1	11
Row 2	6
Row 3	11
Row 4	10
Row 5	8
Row 6	4
Row 7	5

First Survey

goo.gl/vNoeR8

Schedule

Week 1 (Sept 1): Introduction, syllabus, examples, and some basics
Week 2 (Sept 8): Operational palette, structured python, and files
Week 3 (Sept 15): Quantitative plots, plot components
Week 4 (Sept 22): Histograms and distributions
Week 5 (Sept 29): R and ggplot
Week 6 (Oct 6): Images: color, colormaps
Week 7 (Oct 13): Comparisons between datasets
Week 8 (Oct 20): Comparisons between different datasets
Week 9 (Oct 27): Network visualization
Week 10 (Nov 3): Principles of interactive visualization
Week 11 (Nov 10): Interactive visualization with Python
Week 12 (Nov 17): Scientific visualization
Week 13 (Dec 1): Advanced topics
Week 14 (Dec 8): Group presentations

Moved to 213 Greg Hall

Week 3 (Sept 15): Quantitative plots, plot components
Week 4 (Sept 22): Histograms and distributions

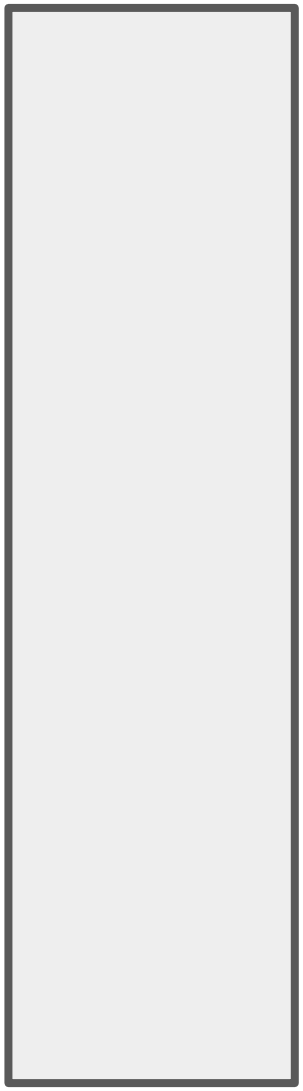
Overview - Themes

1. What are the components of an effective visualization of quantitative data?
2. What tools and ecosystems are available for visualizing data?
3. What systems can be put in place to generate visualizations rapidly and with high-fidelity representation?

Overview - Goals

- Students will be able to communicate information and data through visual representation
- Students will be able to examine a visualization and understand how it can be improved upon
- Students will have facility with the commonplace tools used for visualization, and a deeper understanding of where those tools have shortcomings

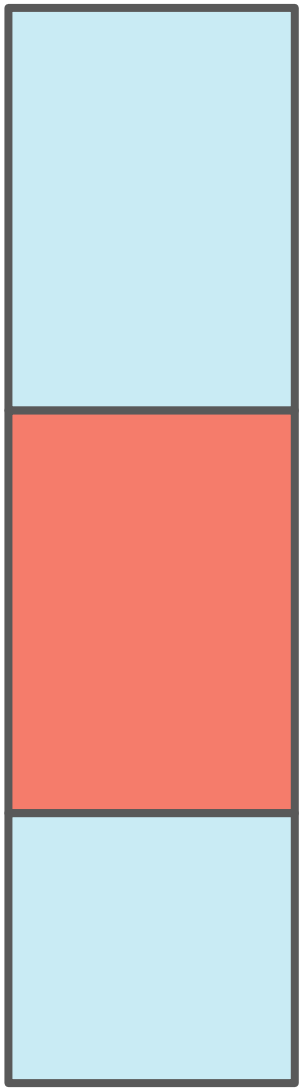
Overview - Structure



1:00 PM

3:50 PM

Overview - Structure



Topic introduction and lecture

Hands-on examples and discussion

Wrap-up

Overview - Grading

70%: Weekly Assignments in prose or code form
30%: Final project

Overview - Assignments

- Weekly, assigned in class, collected following class
- Prose assignments: deconstruction or analysis of a visualization or a dataset.
- Coding assignments: Jupyter notebooks following step by step through collection and processing of data and the visualization of that data

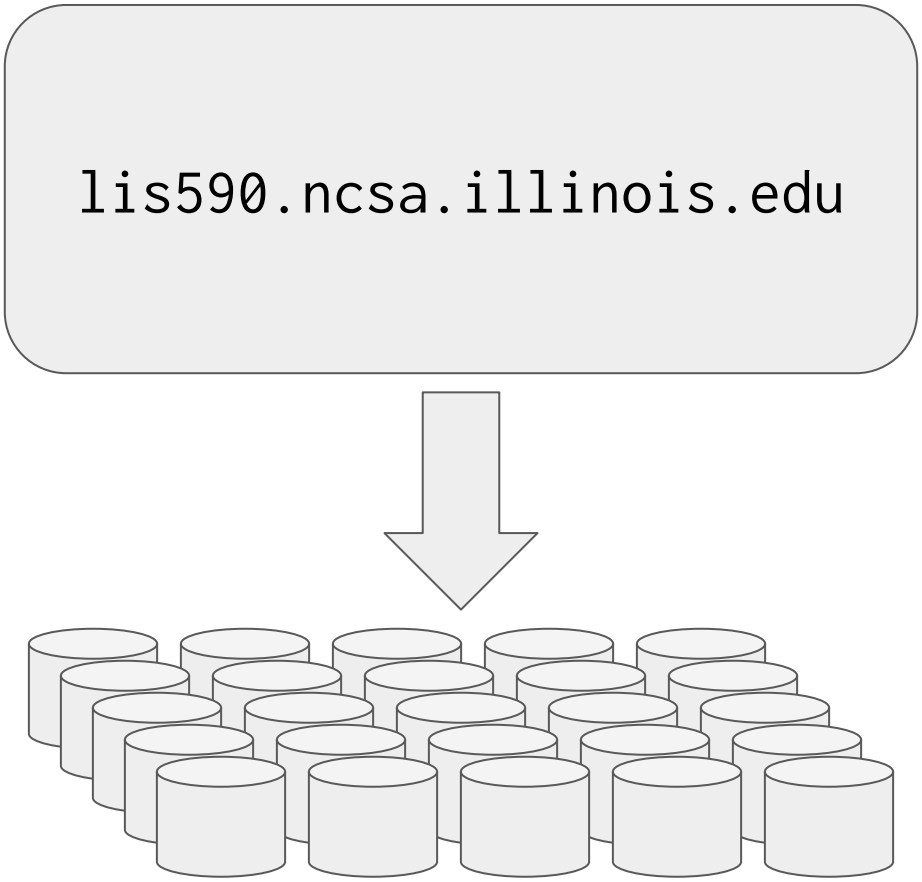
Plagiarism

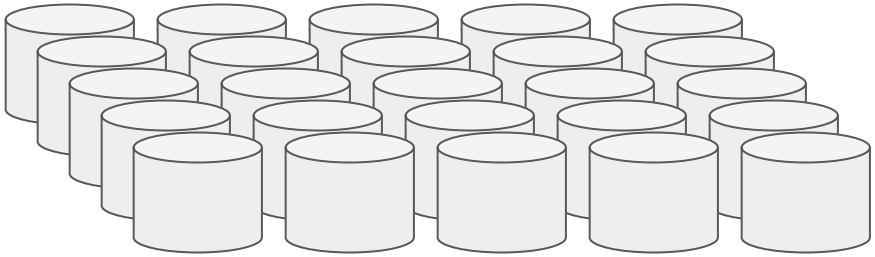
Plagiarism is about copying ideas.

Cite all code you utilize from elsewhere.

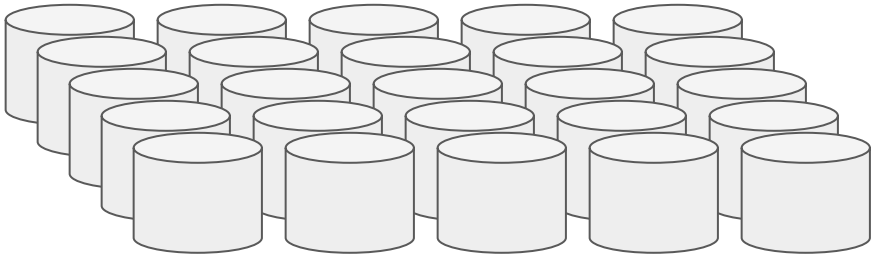
Our Tools

- Python, with some R and Javascript along the way
- Jupyter and Jupyter notebooks on a Jupyterhub, with nbgrader
- The occasional usage of a shell such as bash
- Once in a while some git, and GitHub
- Slack

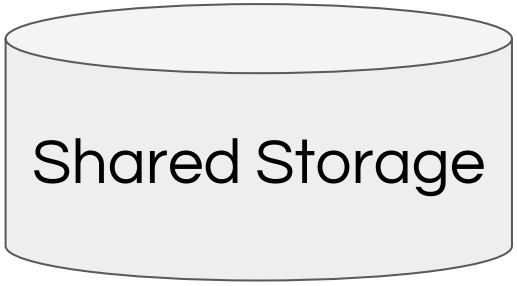

lis590.ncsa.illinois.edu
Jupyter Instances



Jupyter Images



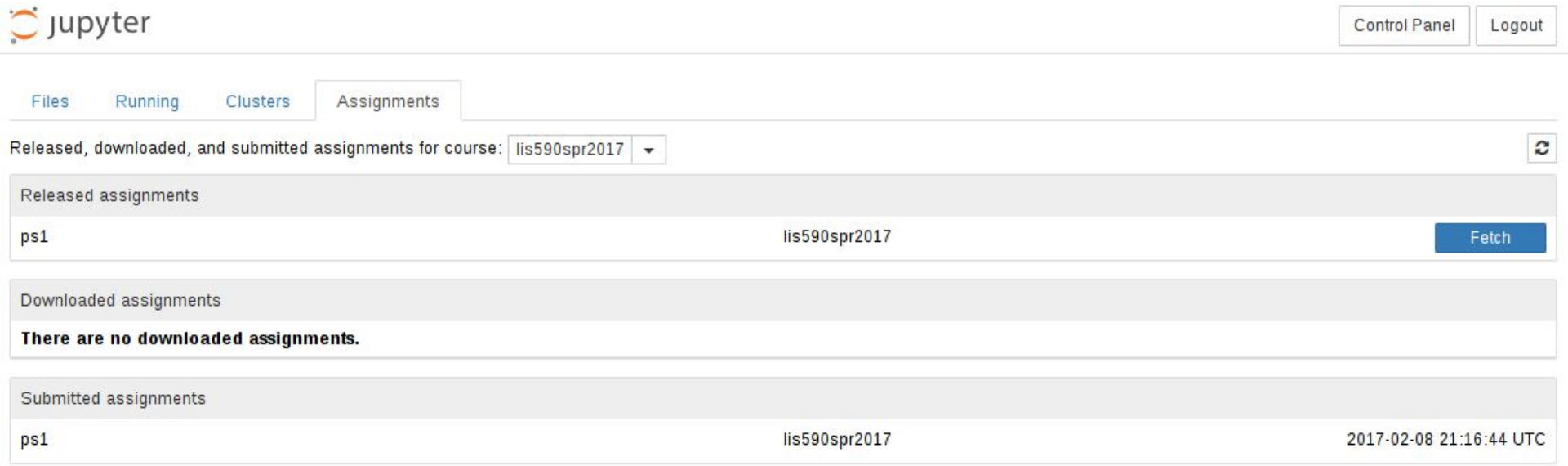
Personal Storage



- Store your notebooks on- and off-site
- Submissions will be via nbgrader
- Data will be available at `/data/` and `/home/$USER/work/data-readonly/`
- Previous lectures will be in `data-readonly/fall2017`
- You will have access to conda, etc, but I may rebuild images to add packages.

Assignment Flow

1. Instructor “releases” an assignment
2. Assignment appears in student “Assignments” tab
3. Students “fetch” assignment, which *copies* it to their work directory
4. Assignments will be notebooks, accessing shared read-only data
5. Upon completion, students “submit” assignment, which *copies* it to the instructor’s inbox
6. Some cells may be “autograded” but feedback will be provided.
7. Once feedback is available, you will be notified.



Slack

- Team is at `lis590dv-fall2017.slack.com`
  - #general : General announcements
  - #assignments : Help with assignments
  - #help : General help with Python, Javascript, visualization, etc
  - #lectures : During lectures, post links, comments, questions here
- Use the @ sign appropriately: @[person], @here, @channel
- Conduct will be held to same standards as any educational venue.
- Web client, standalone client and mobile devices can access this team.
- At the end of the semester, the team will be discontinued.

GitHub

- Reminder: http://github.com/UIUC-iSchool-DataViz/fall2017/
- Lecture notes will be placed there, and available in your JupyterHub instances in data-readonly/fall2017/weekXX .
- Copy the notebooks to your directory before using them.
- If you have not already done so, fill out the form giving me your GitHub usernames.

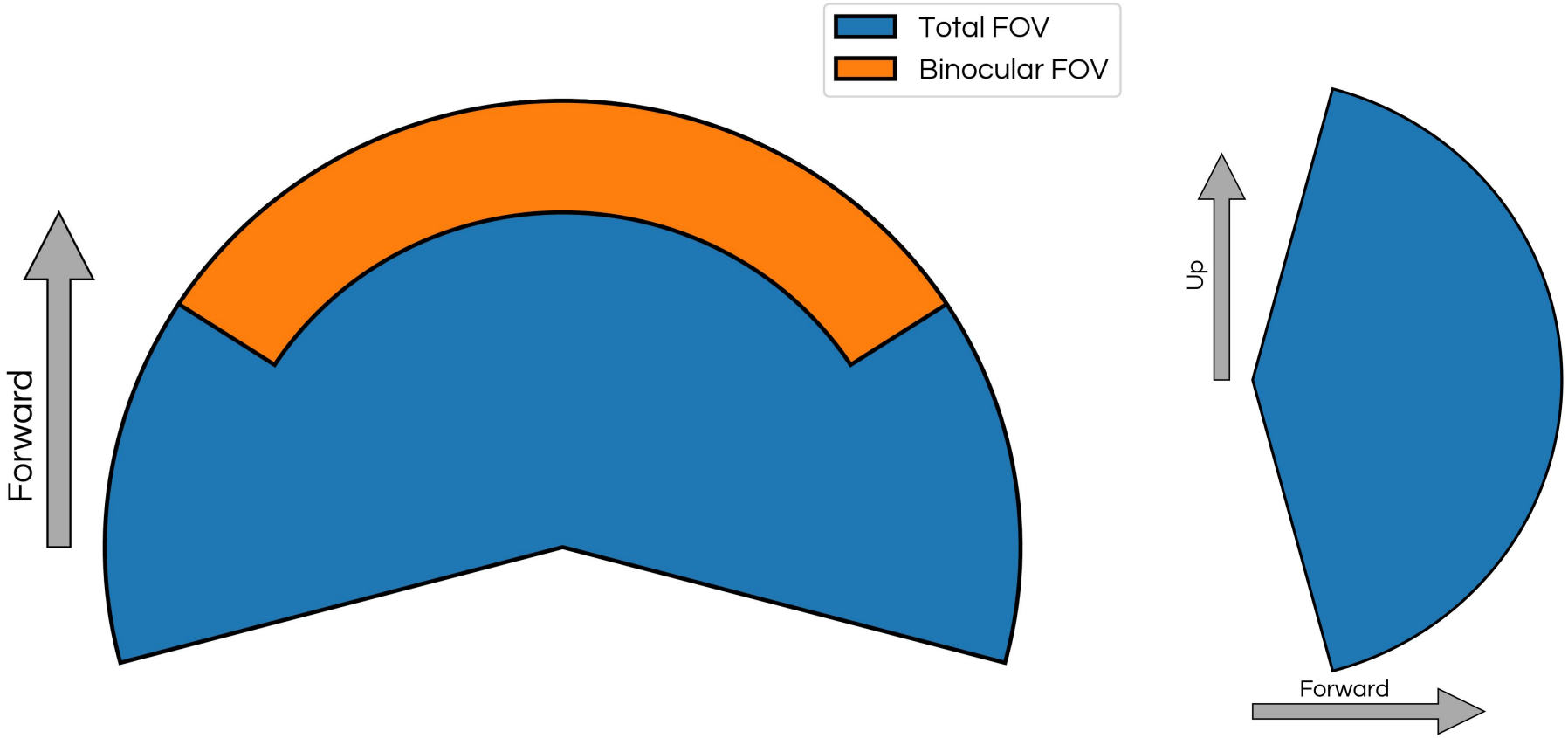
This week:

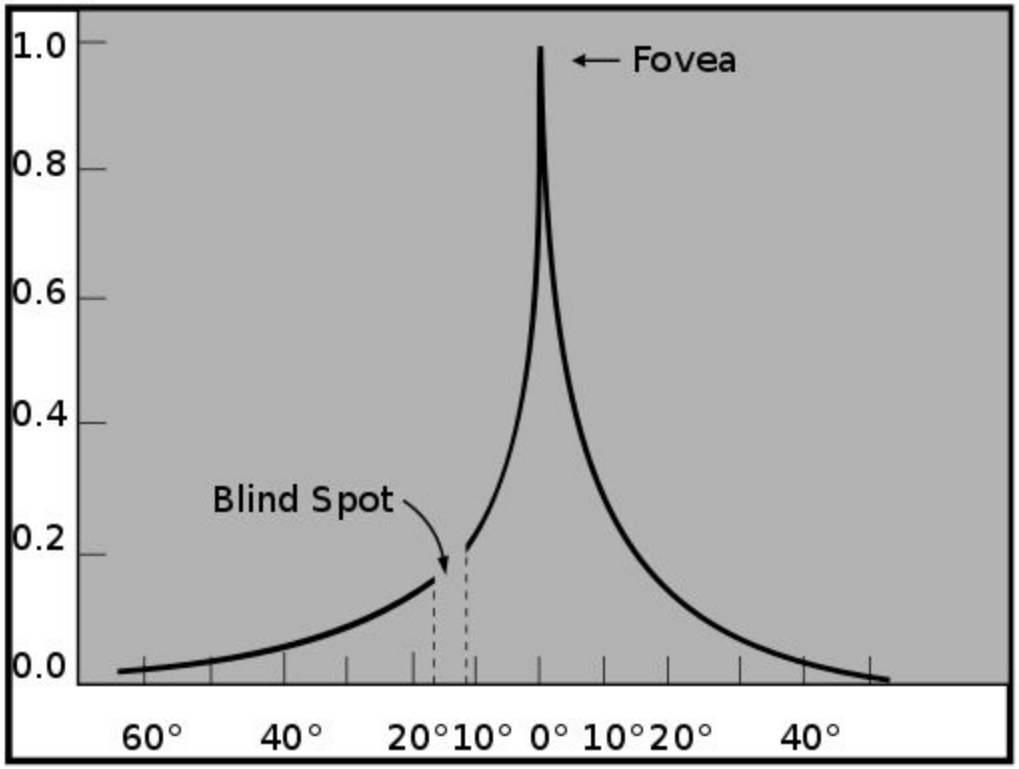
1. Why do we visualize?
2. What types of data do we visualize?
3. How do we visualize?

Why?

(but first, why *wouldn't* we visualize something?)


SONG
peg + cat


Total FOV
Binocular FOV
Forward
Up
Forward



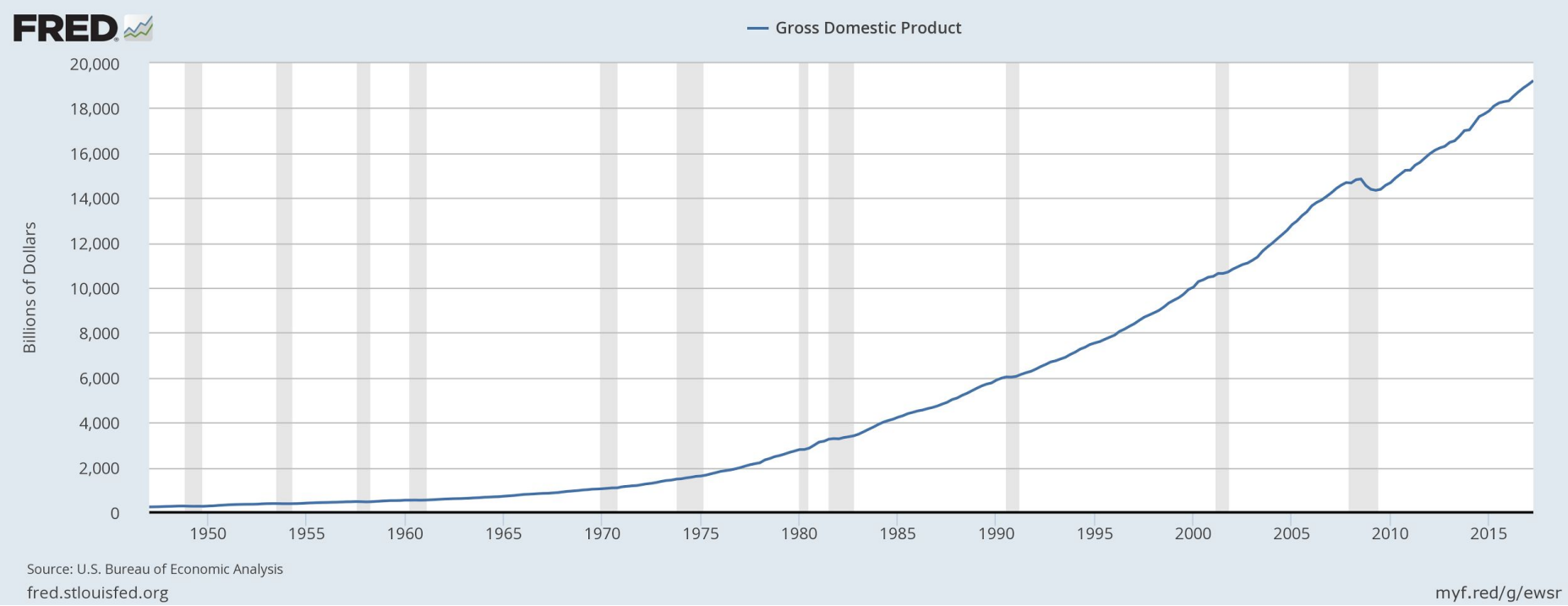
https://commons.wikimedia.org/wiki/File:AcuityHumanEye.svg
By Vanessa Ezekowitz

Now, I need a volunteer!

Read these numbers.

2007-01-01	14233.2
2007-04-01	14422.3
2007-07-01	14569.7
2007-10-01	14685.3
2008-01-01	14668.4
2008-04-01	14813.0
2008-07-01	14843.0
2008-10-01	14549.9
2009-01-01	14383.9
2009-04-01	14340.4
2009-07-01	14384.1
2009-10-01	14566.5

282 records x (?? seconds / 12 records) = ?? seconds

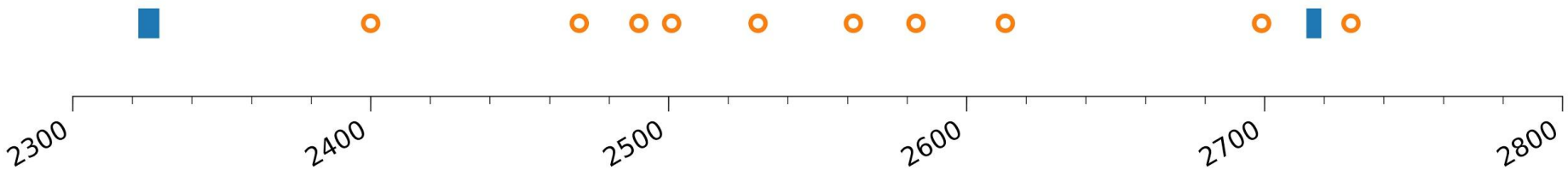


Timeline of Incidents

2729
2714-2719 (Known)
2699
2613
2583
2562
2530
2501
2490
2470
2400
2322-2329 (Known)

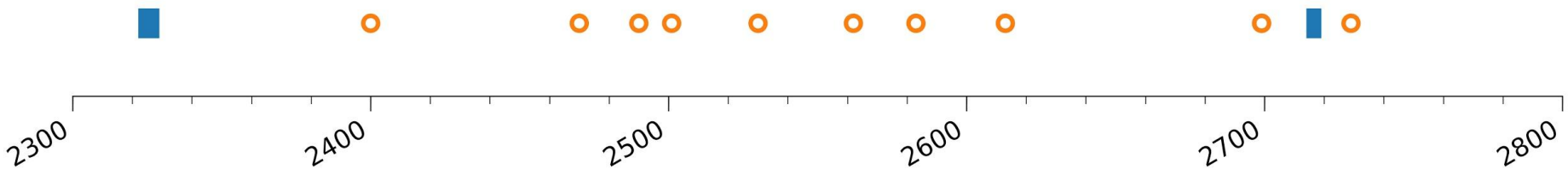
Data from “The Stone Sky” by N.K. Jemisin

Timeline of Incidents



Data from “The Stone Sky” by N.K. Jemisin

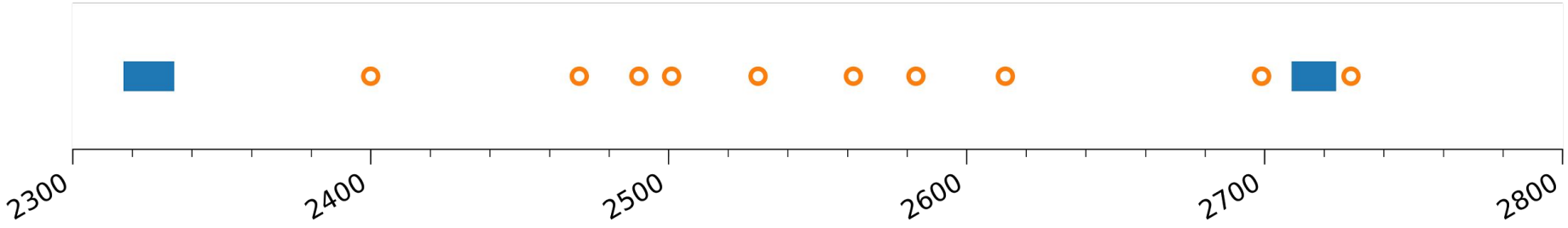
Beware: free parameters!




```
capstyle = “butt”
```

Data from “The Stone Sky” by N.K. Jemisin

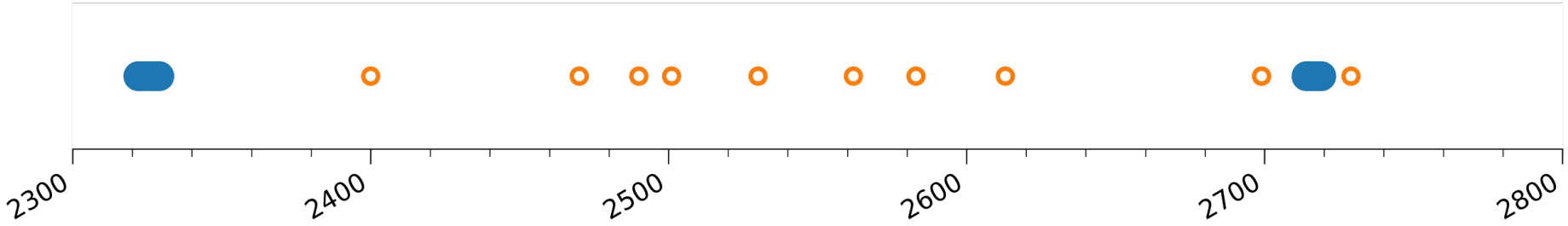
Beware: free parameters!




```
capstyle = “projecting”
```

Data from “The Stone Sky” by N.K. Jemisin

Beware: free parameters!




```
capstyle = “round”
```

Data from “The Stone Sky” by N.K. Jemisin

What?

Visualization for self

Visualization for peers

Visualization for others

How?

“Visualizing data” is not a strict subset of “making an image.”

We tell lies to visualize.

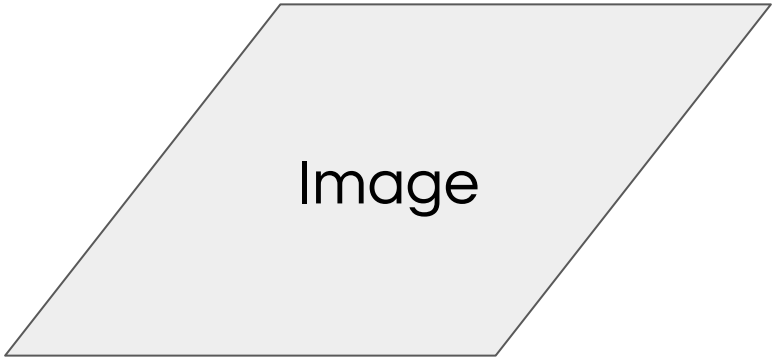
- Collection of the data
- Organization of that data
- Representation of that data


Transformation

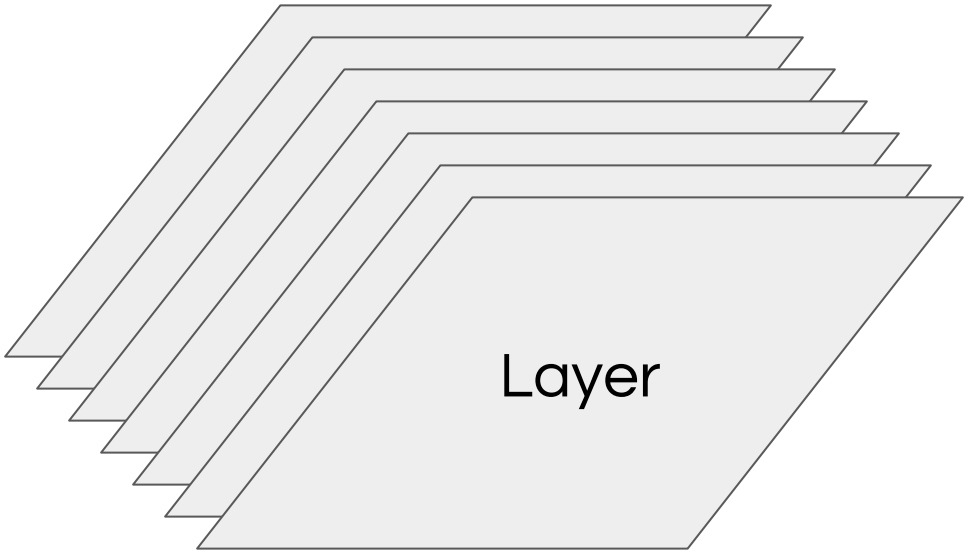
Composition

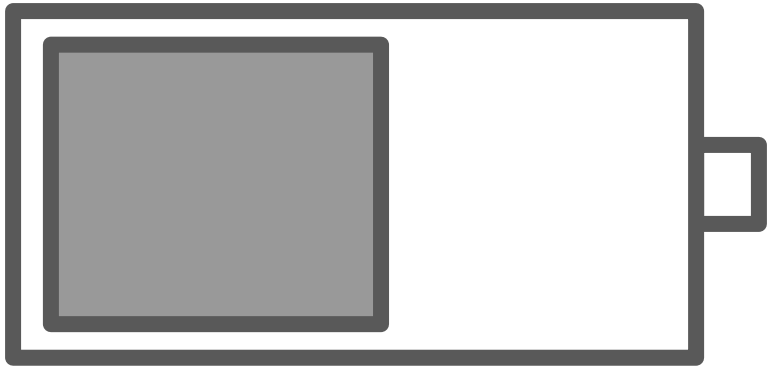

Transformation

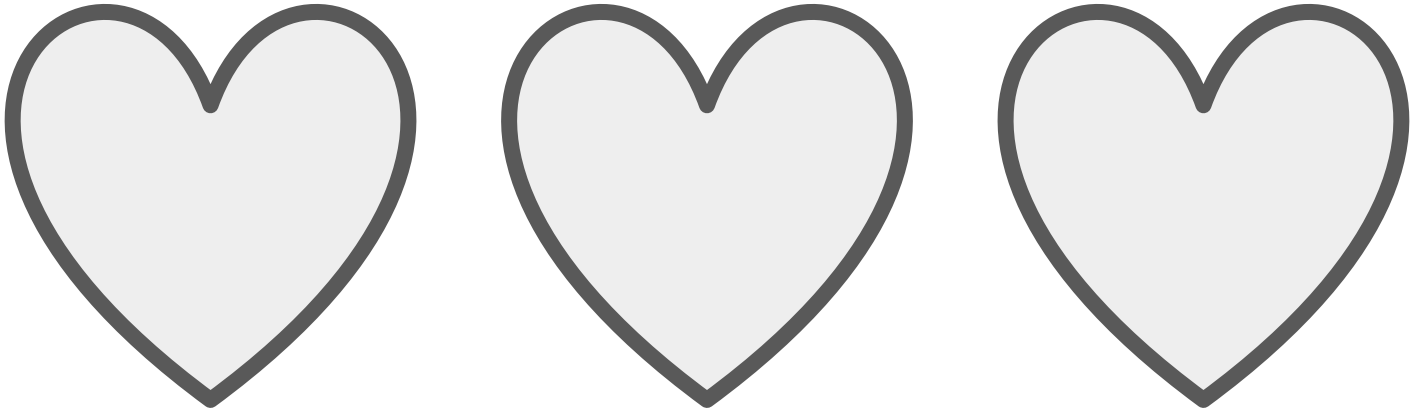
Composition

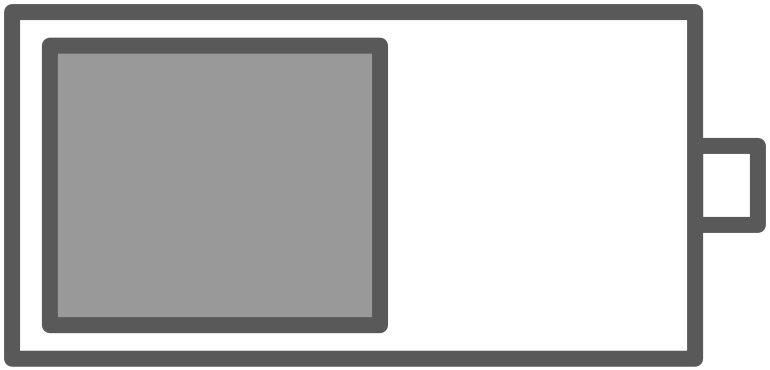


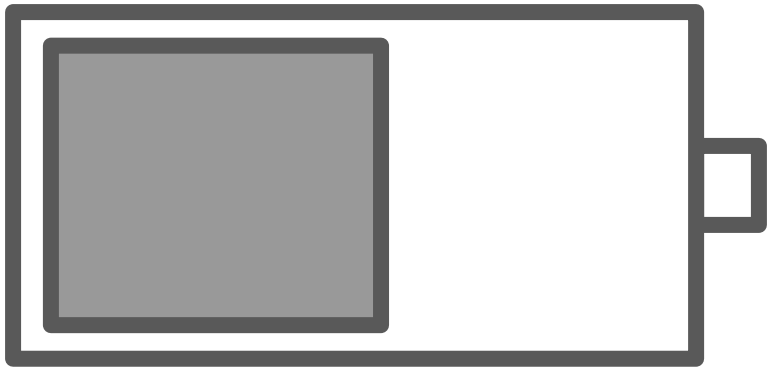
Image


Layer

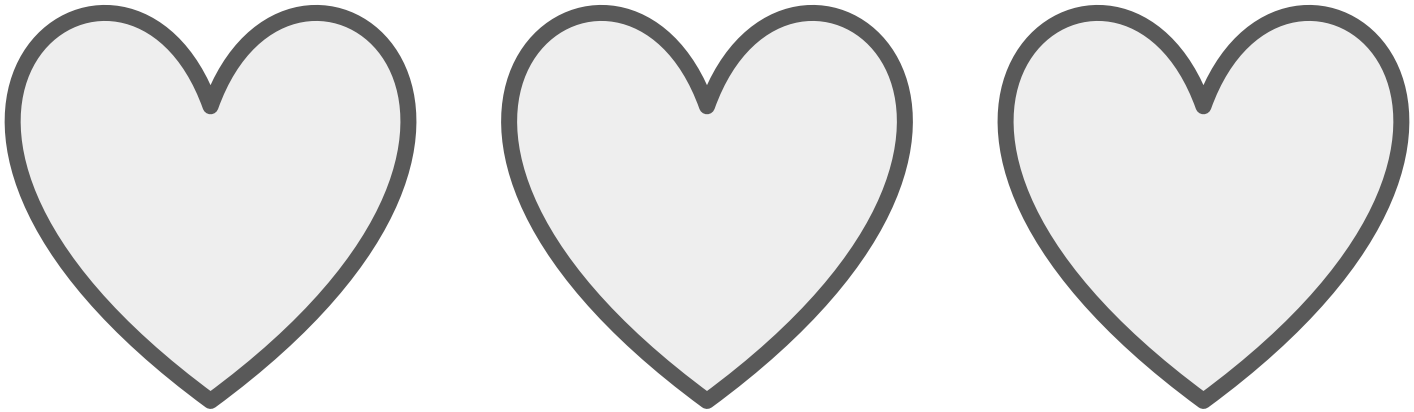




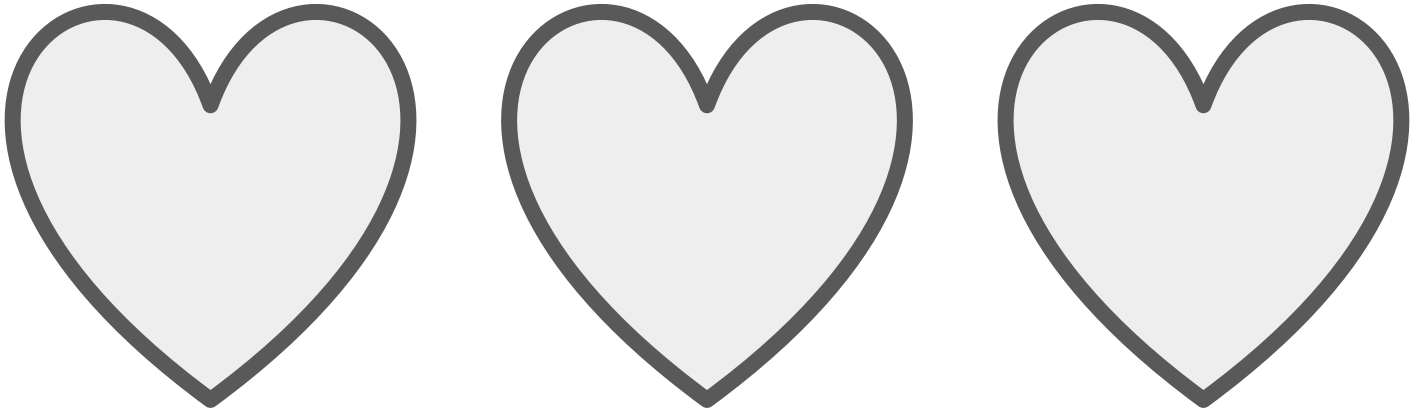




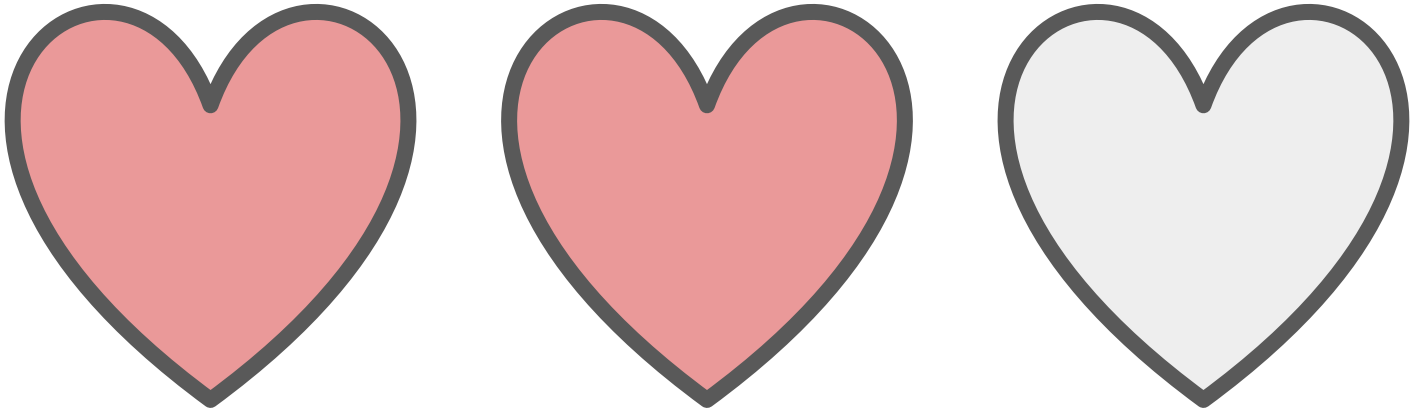
- Sensors read the current “fill” of the battery
  - Analog / digital conversion
  - Normalized with respect to expected “full”
- This is then scaled to a percentage
- The battery image is filled from left to right
- The image is then rasterized and displayed



- Some fixed maximum amount of damage
- Each time damage is taken, decrement
- Each time damage is reversed, increment
- Display number of hearts as appropriate

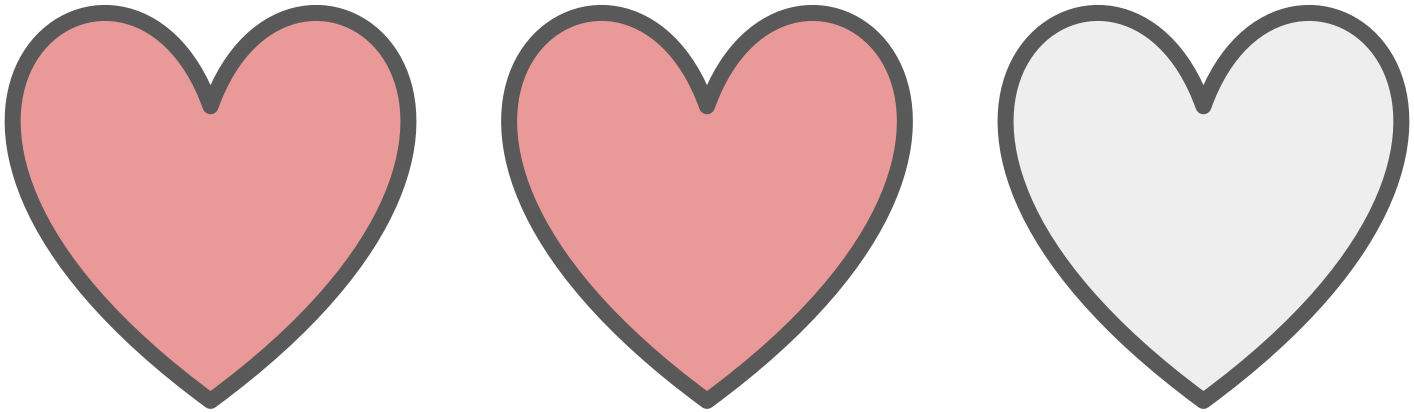


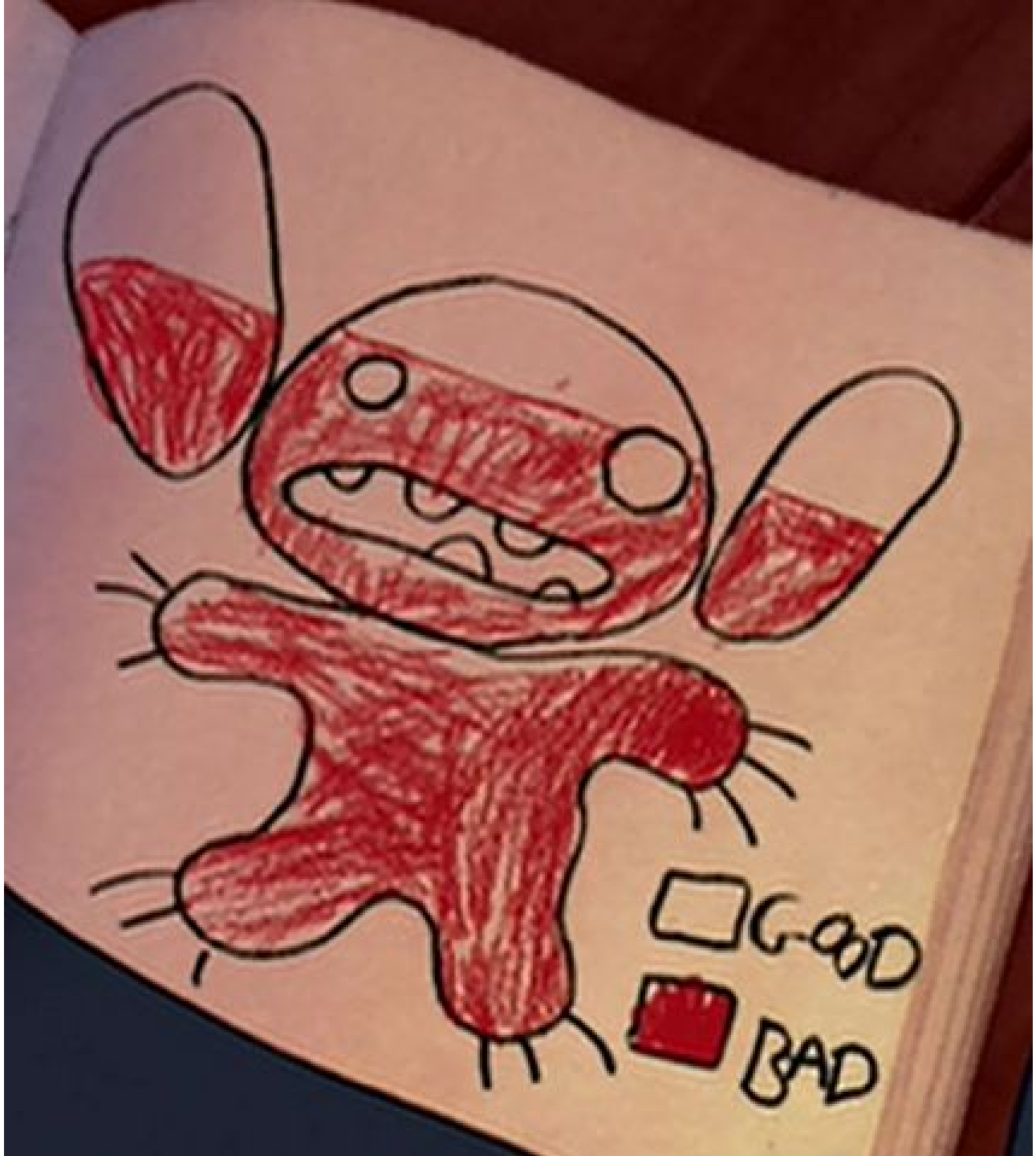
2 of 3 total “points”



Quantitative

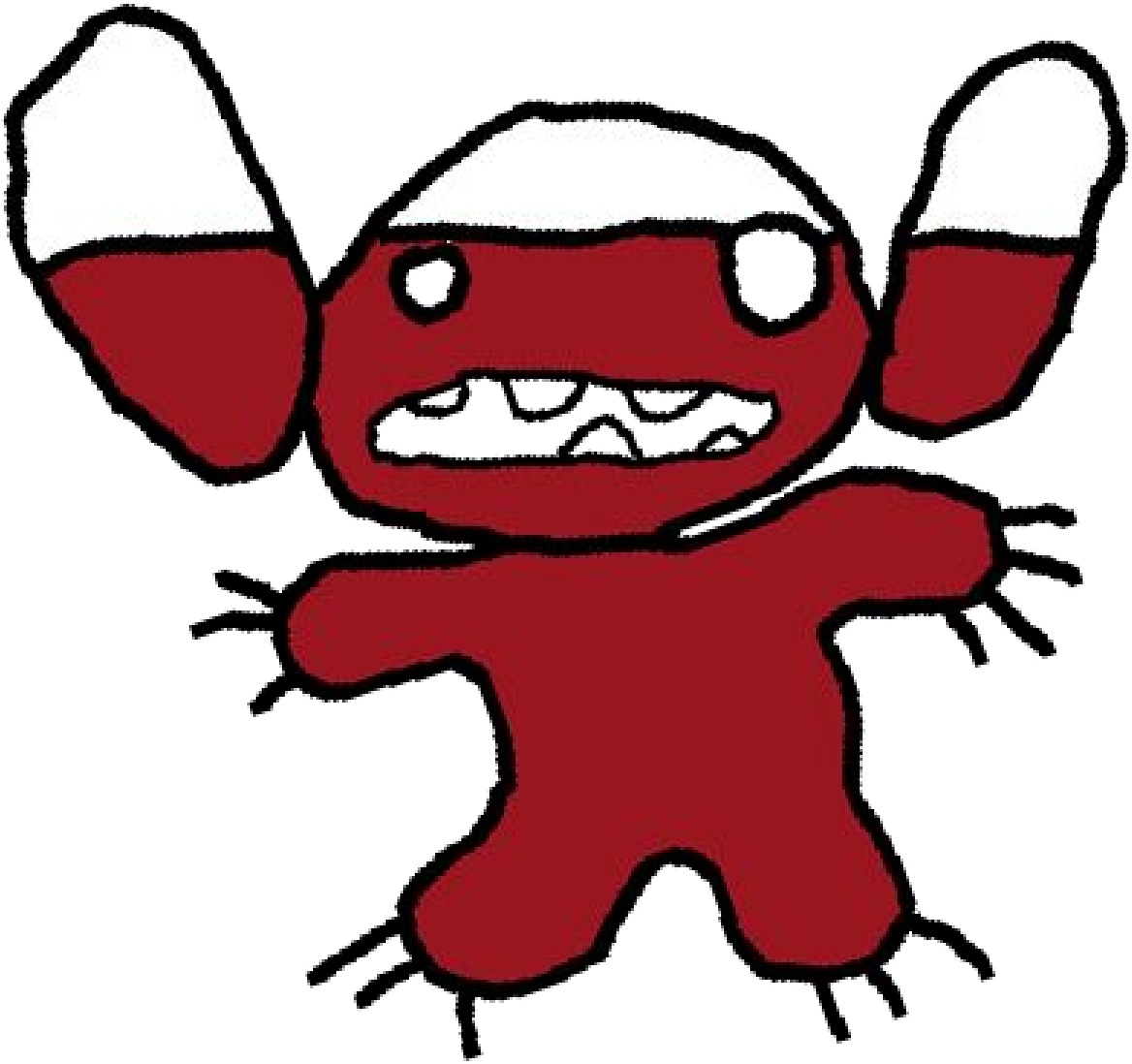
Glyph



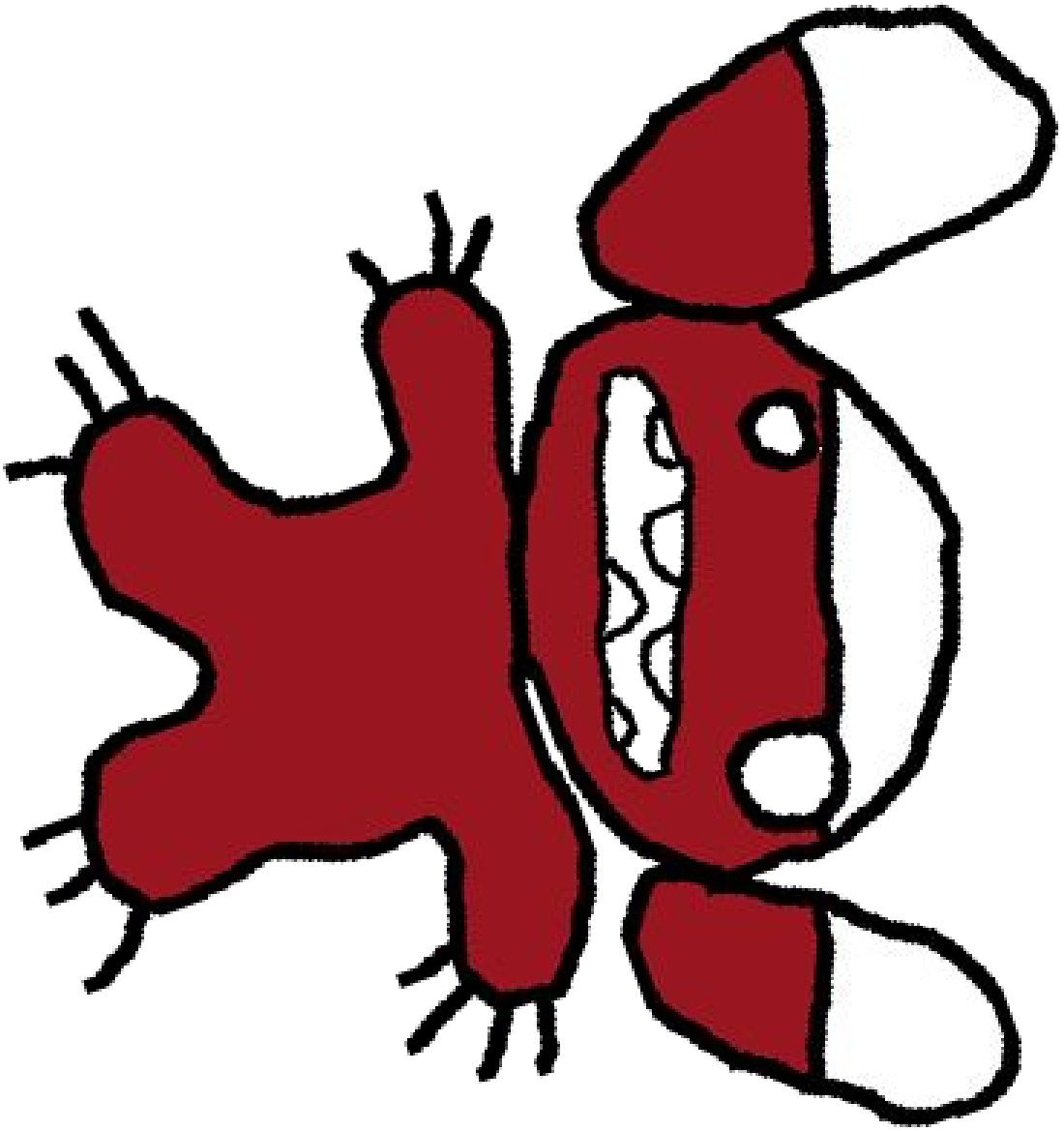

GOOD
BAD

Bad Pixels 40000

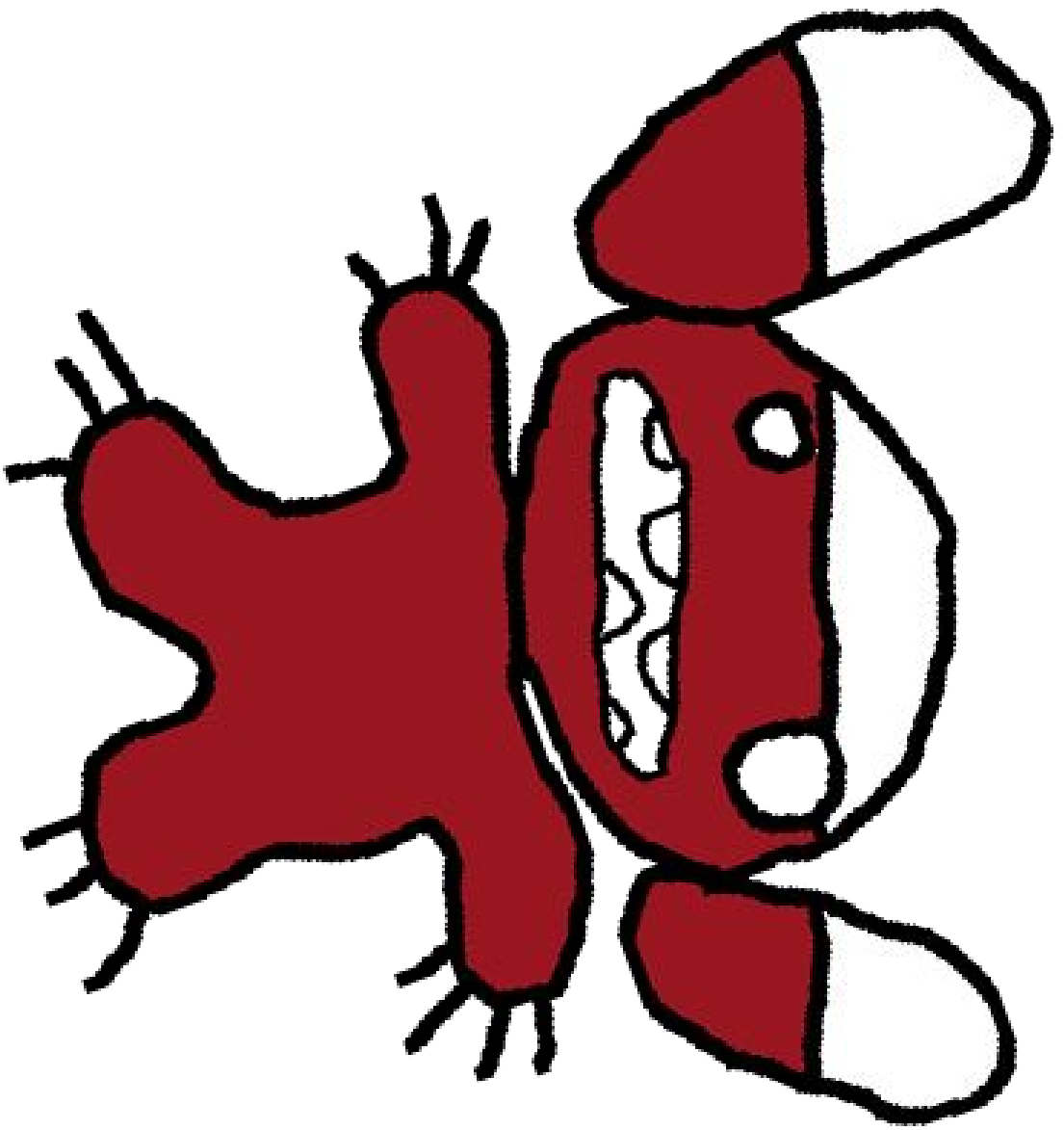
Good Pixels 11500



362 px

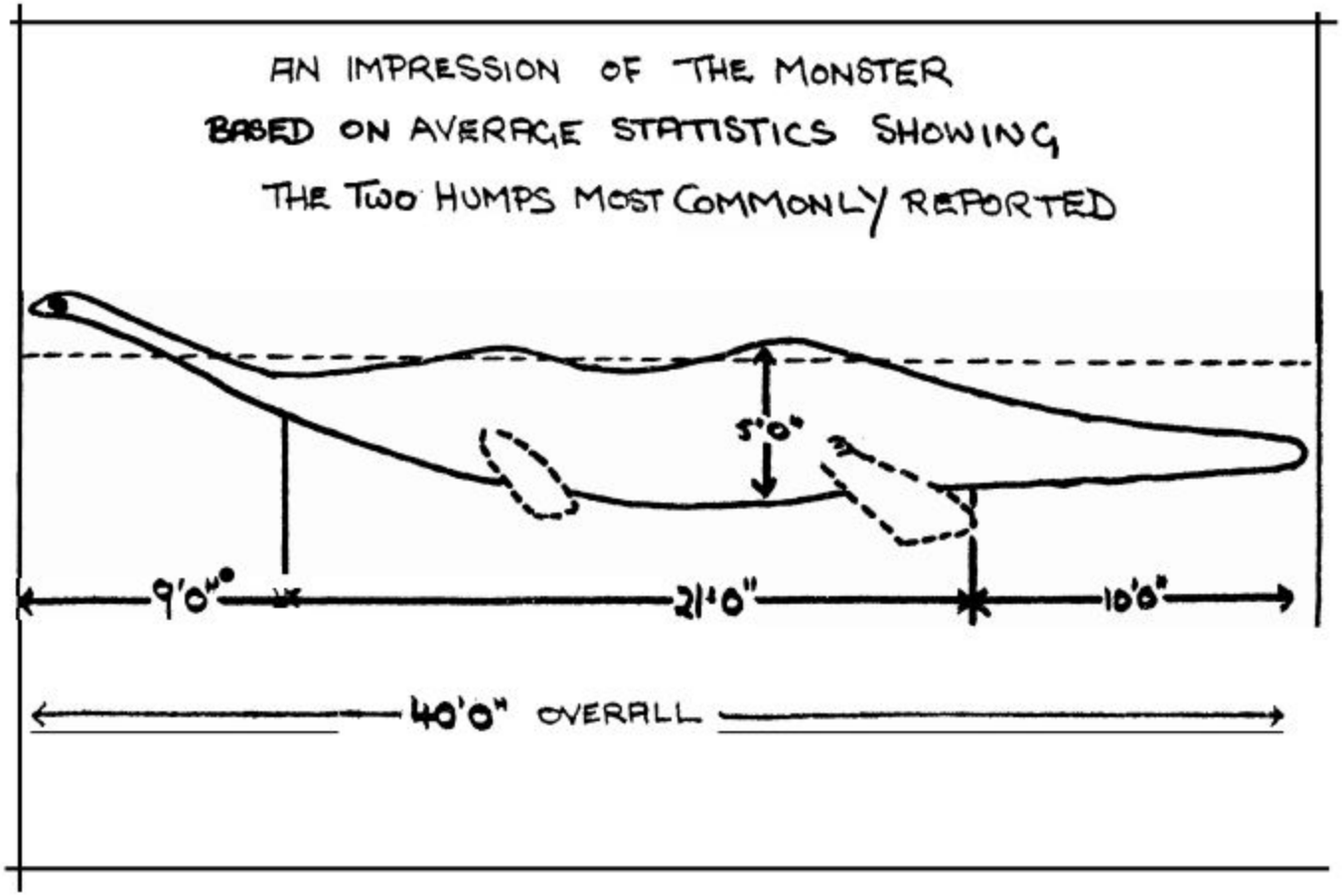


362 px



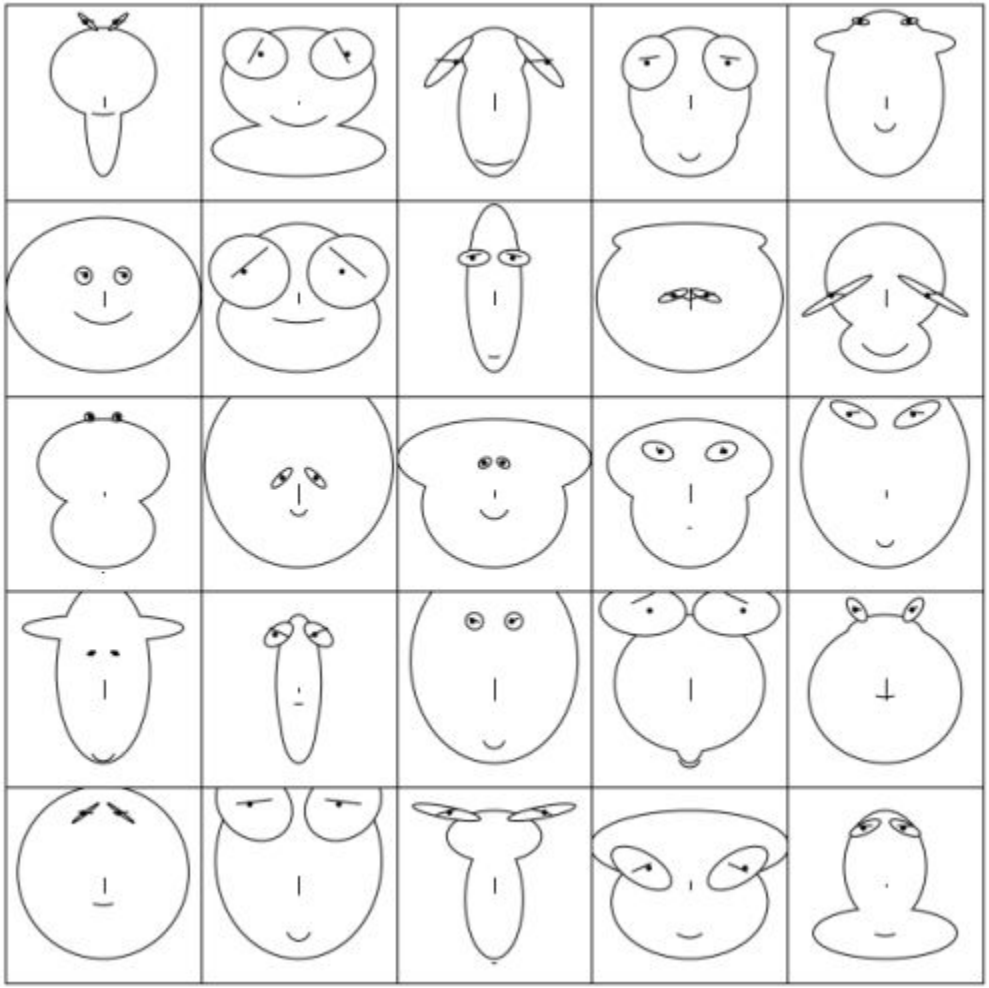
79 px

Linear Scale	78.2%
Area Scale	77.6%


AN IMPRESSION OF THE MONSTER
BASED ON AVERAGE STATISTICS SHOWING
THE TWO HUMPS MOST COMMONLY REPORTED
5'0"
9'0"
21'0"
10'0"
40'0" OVERALL

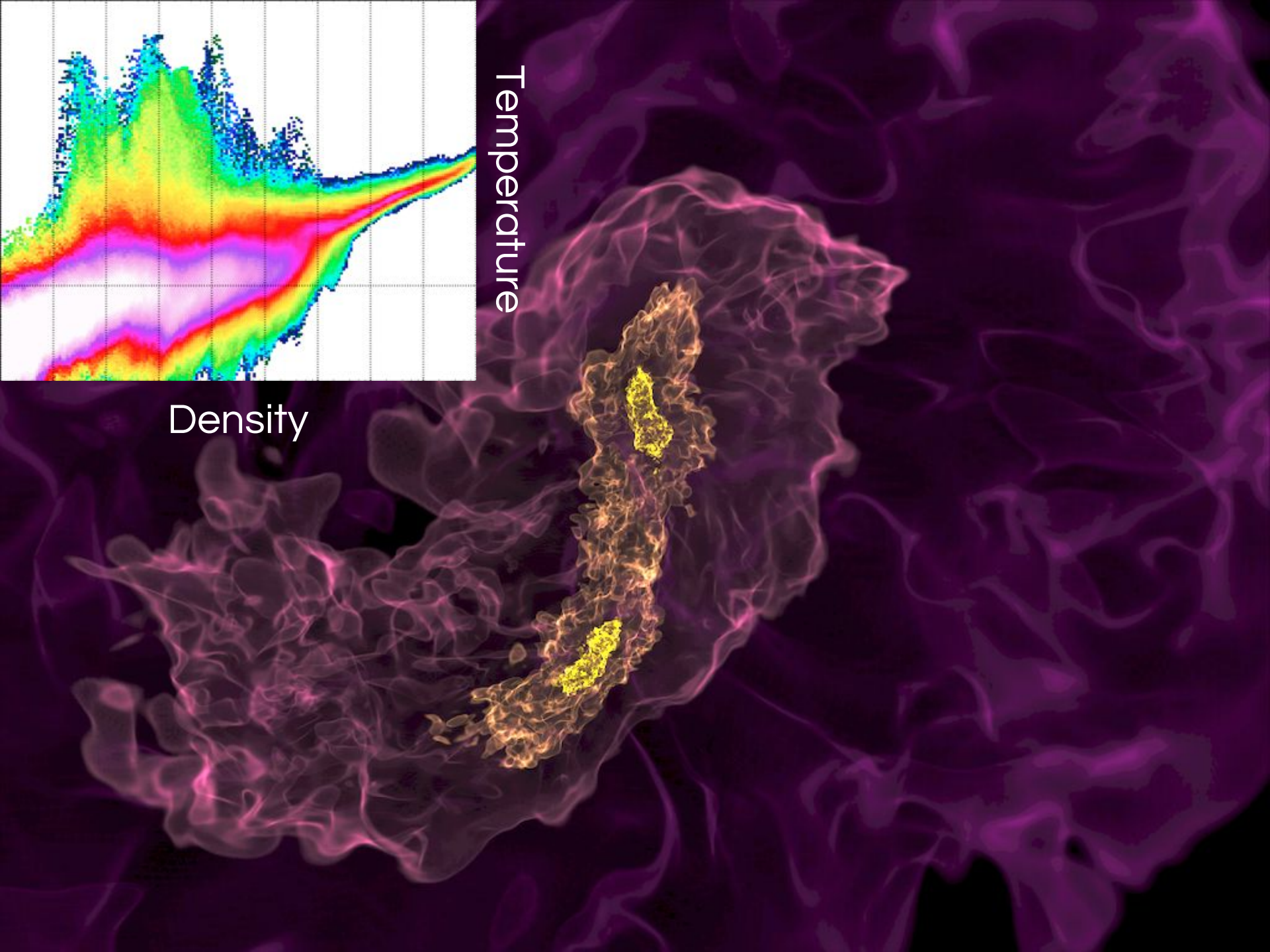
Gowler, 1971

Chernoff Faces



http://www.jstor.org/stable/2284077

https://healthyalgorithms.com/2012/11/12/dataviz-in-python-chernoff-faces-with-matplotlib/


Temperature
Density

Line of sight velocity profiles in turbulent molecular clouds

This is a snapshot from a three dimensional simulation of about 5,700 Solar masses of gas in a 10 pc periodic box. The gas began with fully developed turbulence at Mach 8; this is after 1.27 Myr of self-gravitating evolution. Click around the image to explore the line-of-sight velocity structure.

*pick a column density image (current choice is **darker**)*

total column density (log scale)

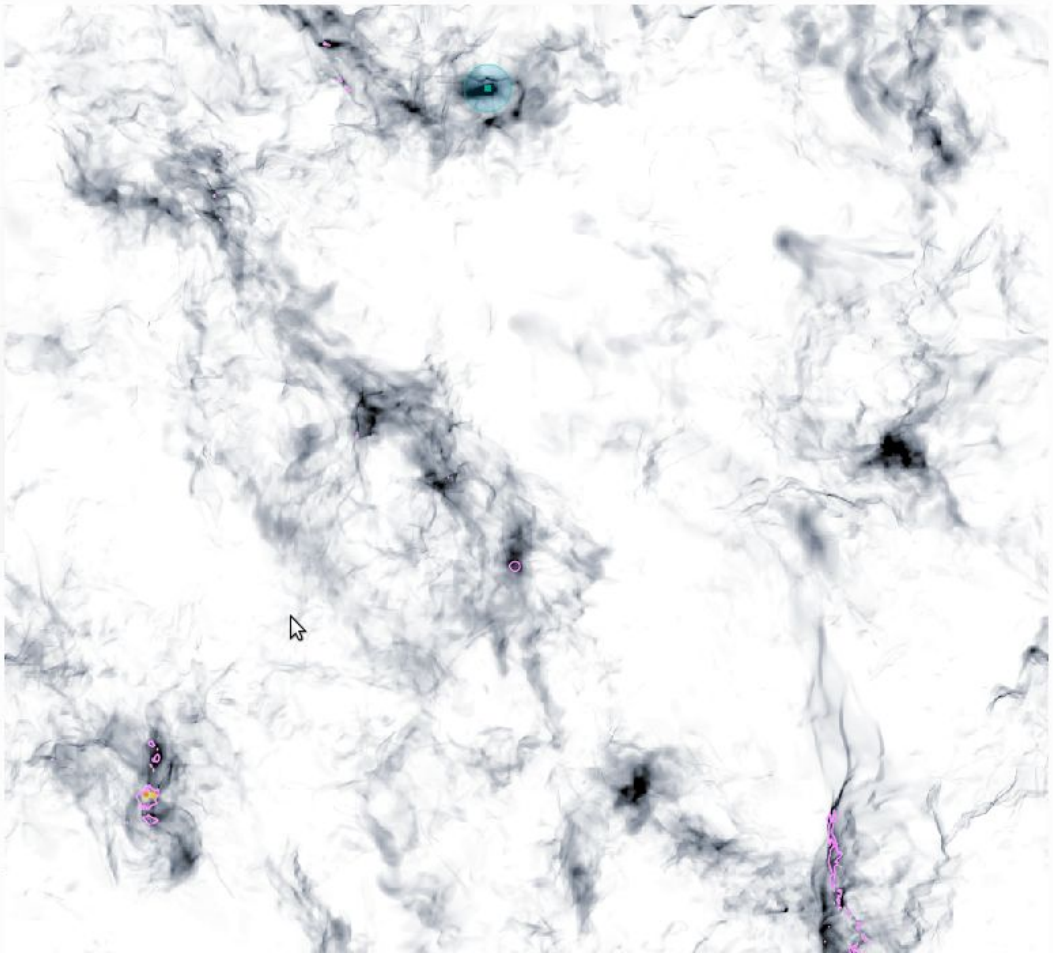
10^3 < n cm^{-3} < $10^{4.5}$ (linear scale; approximates C^{18}O)

toggle high density contours and sink particles

$10^{4.5}$ < n cm^{-3} (approximates N_2H^+) **sink particles**

click the image to inspect a point

tracer :	total	C^{18}O	N_2H^+
log N / cm^{-2} :	22.26	22.18	0

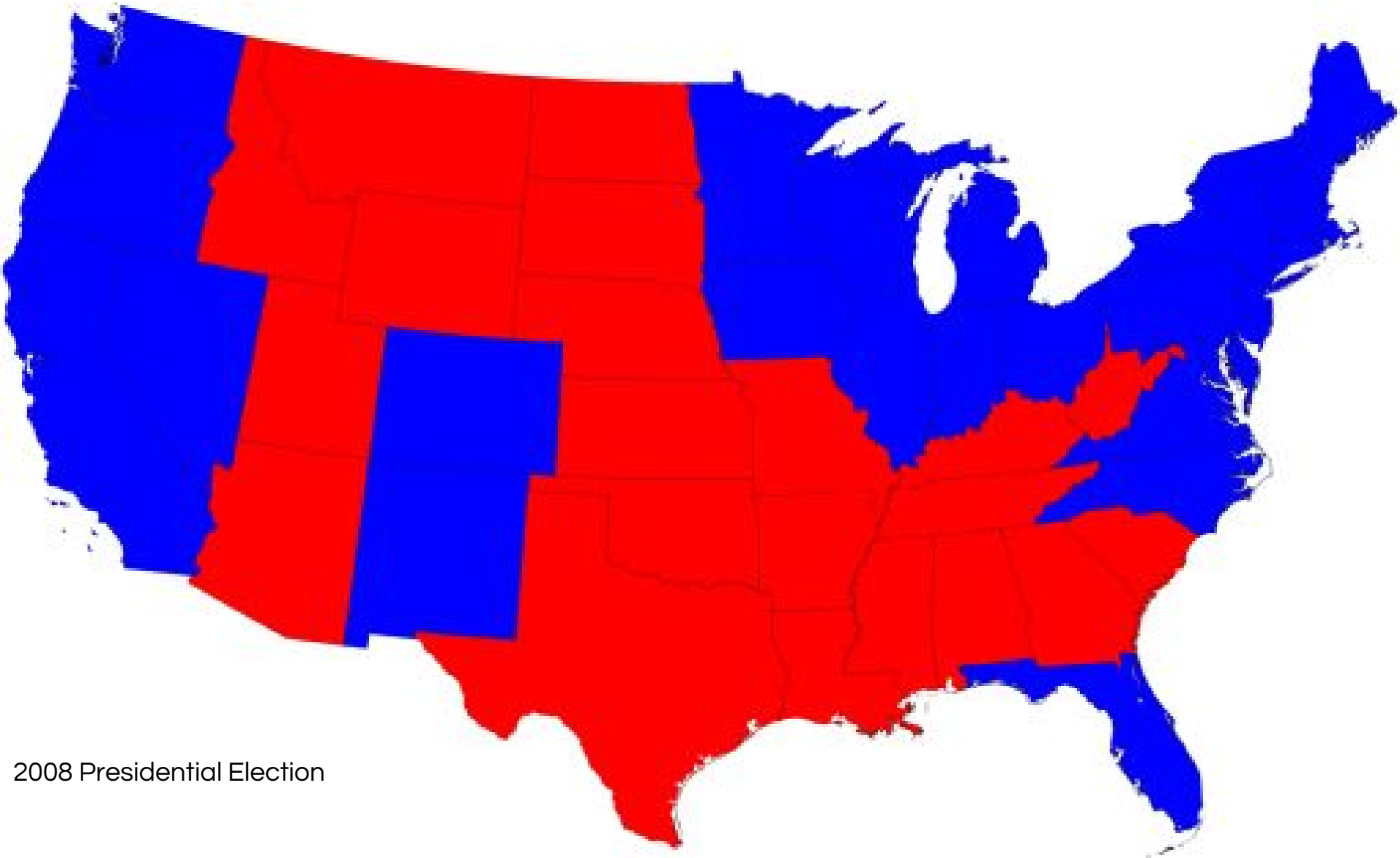


We tell lies to visualize,
but we must be honest.

Our choices must be

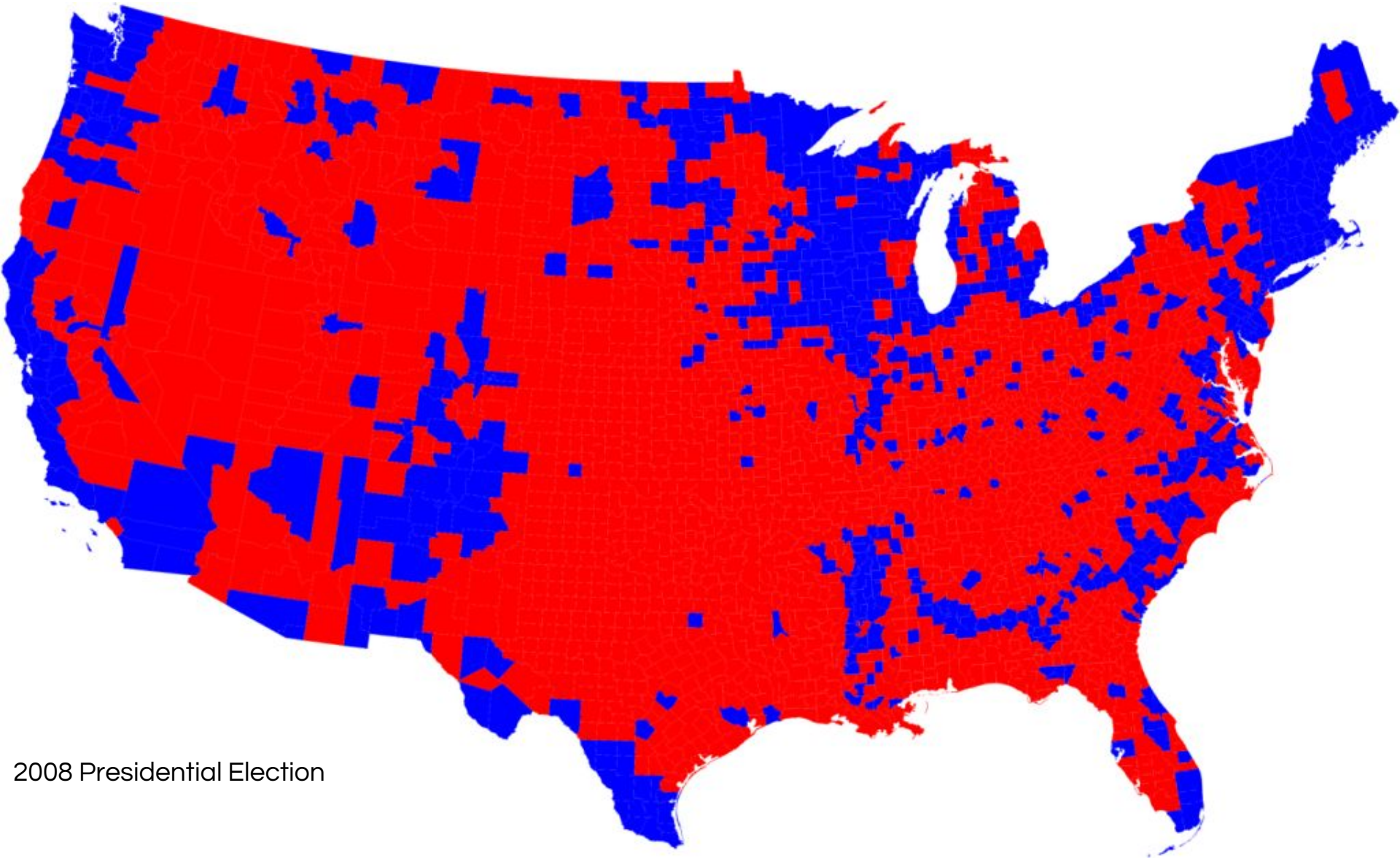
- Deliberate
- Informed
- Motivated
- Justifiable

Source: Mark Newman, University of Michigan
http://www-personal.umich.edu/~mejn/election/2008/



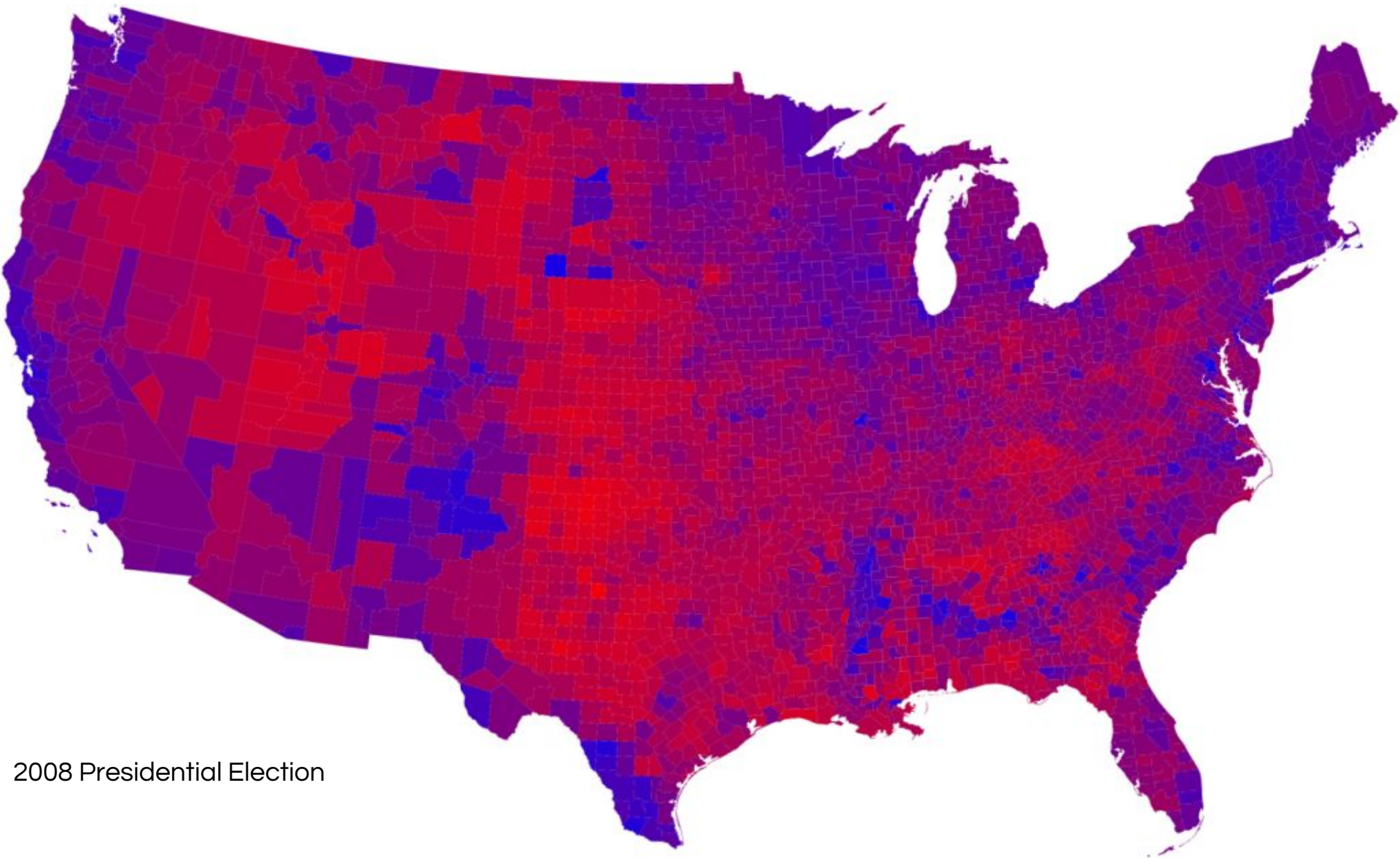
2008 Presidential Election

Source: Mark Newman, University of Michigan
http://www-personal.umich.edu/~mejn/election/2008/



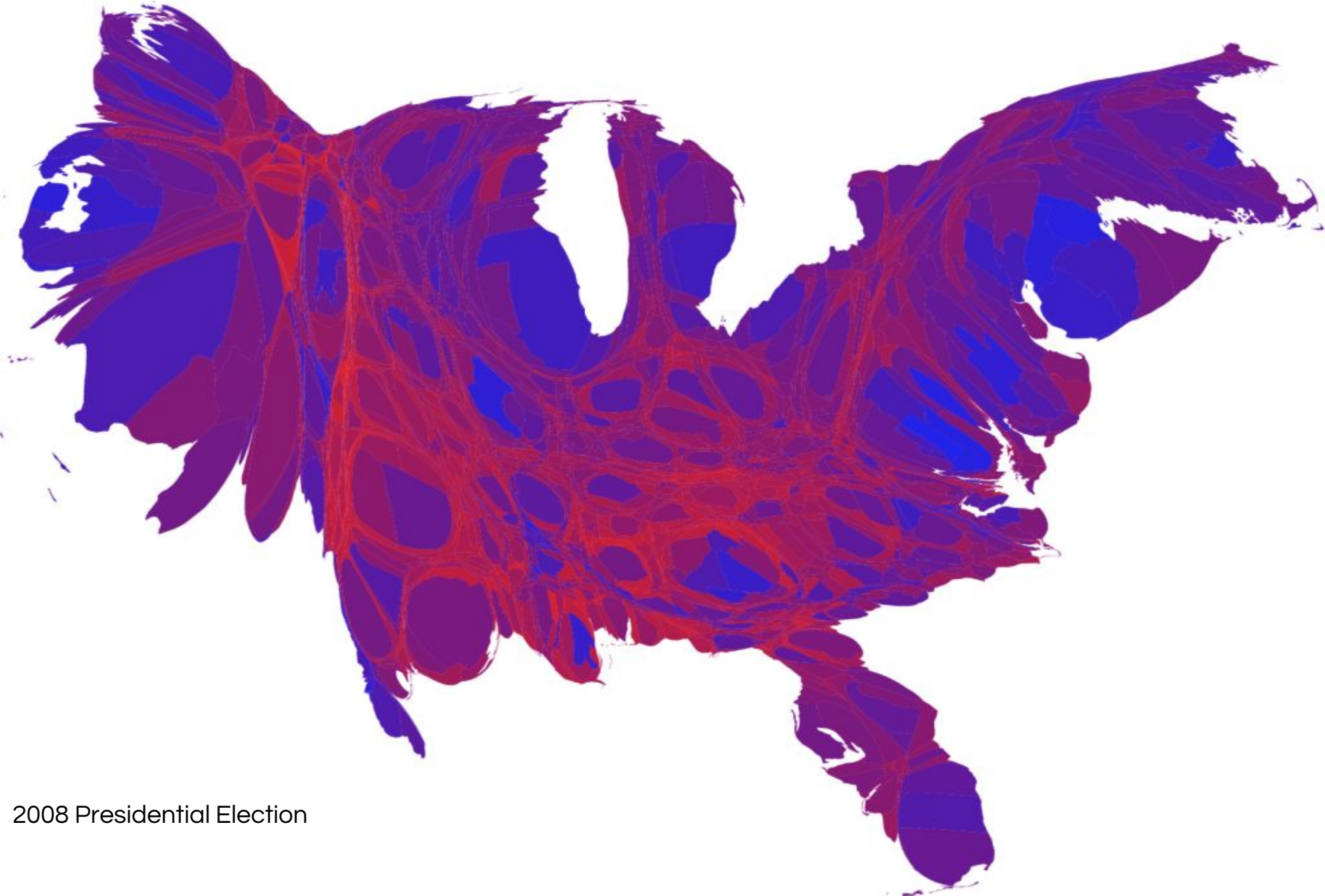
2008 Presidential Election

Source: Mark Newman, University of Michigan
http://www-personal.umich.edu/~mejn/election/2008/



2008 Presidential Election

Source: Mark Newman, University of Michigan
http://www-personal.umich.edu/~mejn/election/2008/



2008 Presidential Election

This week:

1. We visualize to change how we understand things.
2. We visualize data for ourselves, for our peers, and for others.
3. Visualization is a series of steps that we take to produce a different representation of data.

Assignment 1

- Identify three visualizations in pop culture
  - Movies / TV / Music videos
  - Everyday life
  - Advertisements
- Describe each one in detail
  - Where did the data come from?
  - Is the data quantitative, qualitative, categorical, etc?
  - How was the data processed before being displayed?
  - What method was used to display that data?
- Replicate the visualization with different, but similarly “shaped,” data
  - By hand is acceptable
  - Computational methods should include source code