Data Visualization

Drawn from Chapter 1 of **Beautiful Visualization**

Why should we visualize?

What is the goal of visualizing information?

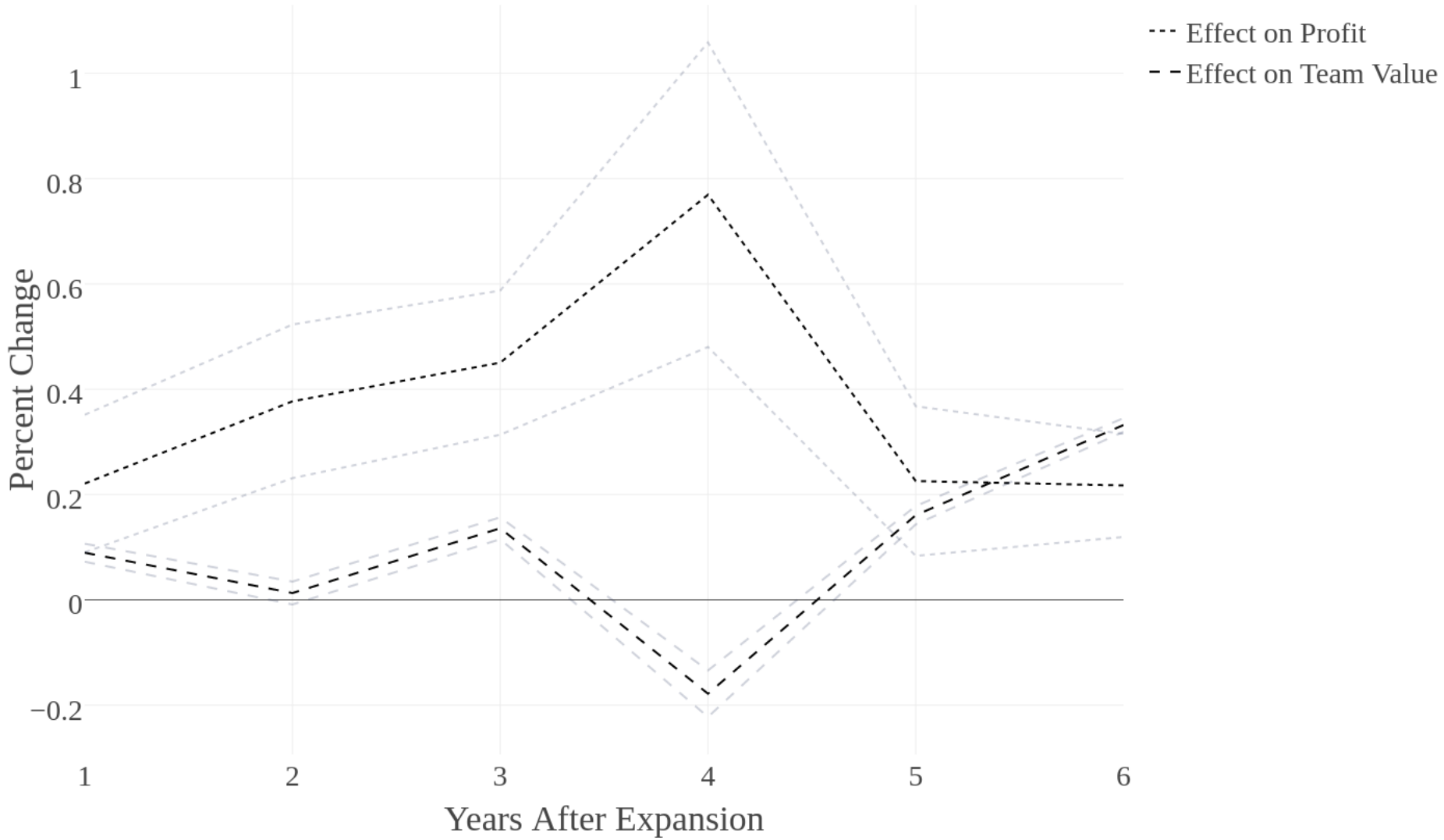
- We are translating information from an effective method for **storing** it to an effective way of **consuming** it!
- My brain (and probably yours, too!) does not like to consume spreadhseets
- We perceive the world visually!

Table 3: Fixed Effects Model results for NFL Panel Data

Variable	Effect on Profit	Effect on Team Value
Intercept	-0.273	6.267
	(0.596)	(0.068)
Expansion Past Year	0.221***	0.089***
	(0.066)	(0.009)
Expansion (2 Years Ago)	0.377***	0.013
	(0.074)	(0.011)
Expansion (3 Years Ago)	0.45***	0.136***
	(0.07)	(0.011)
Expansion (4 Years Ago)	0.769***	-0.178***
	(0.147)	(0.022)
Expansion (5 Years Ago)	0.226***	0.161***
	(0.072)	(0.009)
Expansion (6 Years Ago)	0.217***	0.332***
	(0.05)	(0.007)
TVDeal	0.22**	-0.04**
	(0.108)	(0.016)
Labor Contract Past Year	-0.272***	0.5***
	(0.095)	(0.012)
Playoffs Past Year	0.011	0.008
	(0.061)	(0.008)
Super Bowl Past Year	0.049	0.036**
	(0.1)	(0.015)
Revenues	0.017	0.0***
	(0.003)	(0.0)
% Change in Team Value	0.006	0.004***
	(0.004)	(0.001)
N	509	537
Adj. R^2	.508	.967

Note: * Significant at 10% Level, ** Significant at 5% Level, *** Significant at 1% Level
Both regressions also include team effects, year effects, and profit and team value (where they are not the dependent variable).
Results are robust to excluding the alternate dependent variable from the X vector.

A rockin' table


Effect on Profit
Effect on Team Value
Percent Change
1
0.8
0.6
0.4
0.2
0
−0.2
1
2
3
4
5
6
Years After Expansion

Data is the art of storytelling, so let's really make it ART

When we present data, we are trying to tell a story.

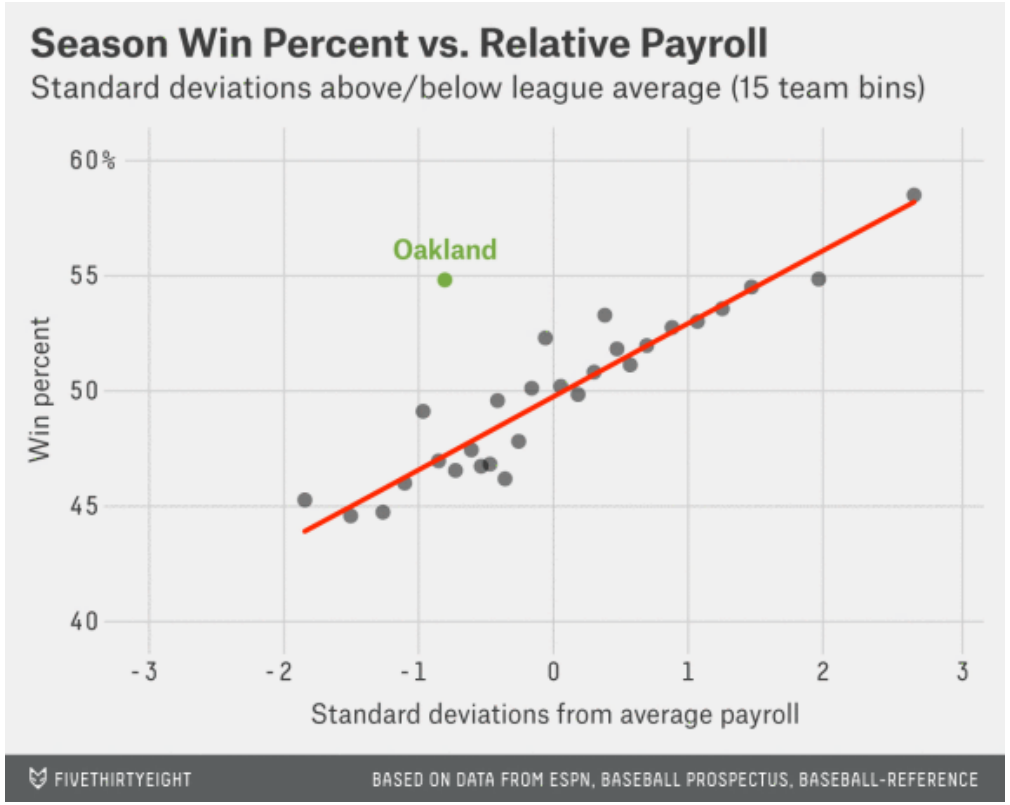
- Kids prefer stories with illustrations, because they are only just learning about how the world looks and works
- Visualization aids unfamiliar audiences

Find a clear story to tell, and let your visuals help you tell it.

What are the challenges of visualization?

1. Dimensionality
    - We often have a lot of different features in our data
    - We can’t easily process more than 2 or 3 at once
2. Context
    - It can be hard to understand a figure’s context

Classic Visuals



Classic Visuals

Our visuals should be

1. Aesthetically Pleasing
2. Novel
3. Informative
4. Efficient

Aesthetically Pleasing

- Don’t let beauty overwhelm data
- Aesthetics should accentuate the information
- Familiar looks and feels can help!

Novel

A visual can be novel in many ways:

- Novel Data
- Novel Insights
- Novel Presentation

> Most often, designs that delight us do so not because they were designed to be novel, but because they were designed to be effective -- Beautiful Visualization

Informative

> A visual that [is not informative] has failed. -- Beautiful Visualization

Ask: What is the intended usage of our visual?

> My goal is to display _____ so that _____ can _____.

- What is our context of use?
- Is it for presentation or exploration?

Catering a visual to our audience ensures that they can quickly obtain the most valuable information.

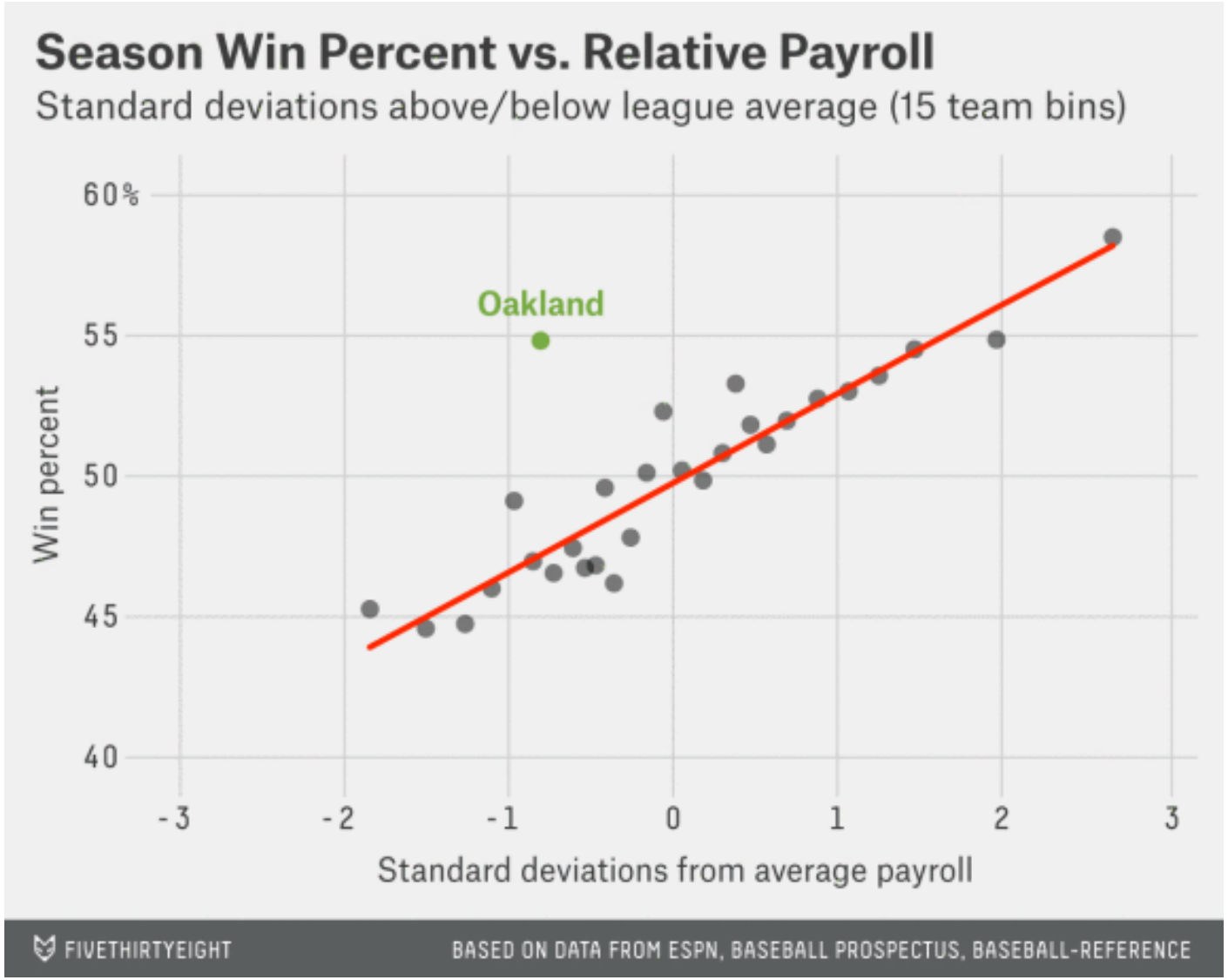
Informative

Efficient

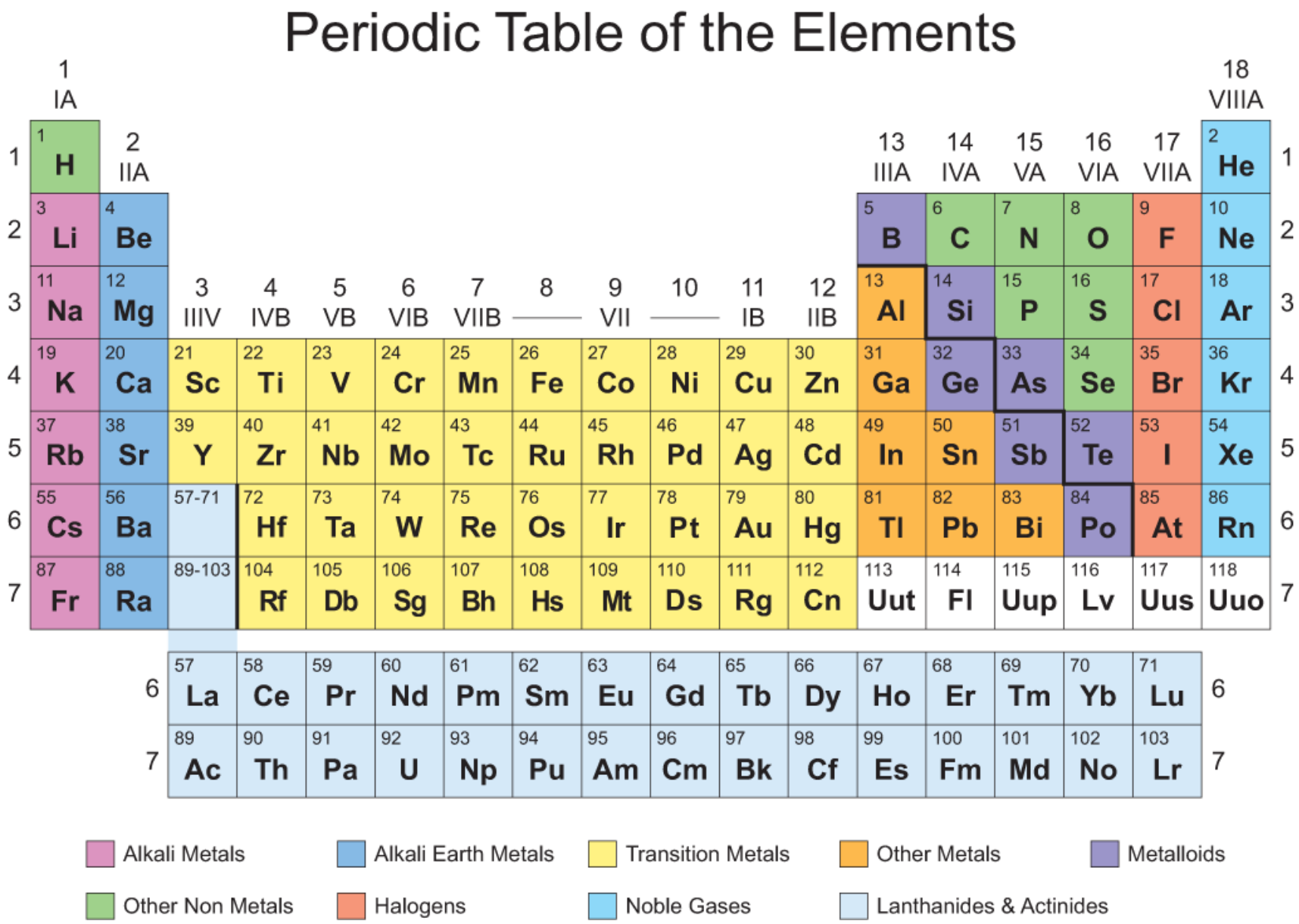
> Irrelevant data is the same thing as noise. If it’s not helping, it’s probably getting in the way. -- Beautiful Visualization

- The minimum viable product concept is critical in visualization of data
  - Each new element slows your audience's perception of the important points
  - BUT! Don't omit critical components

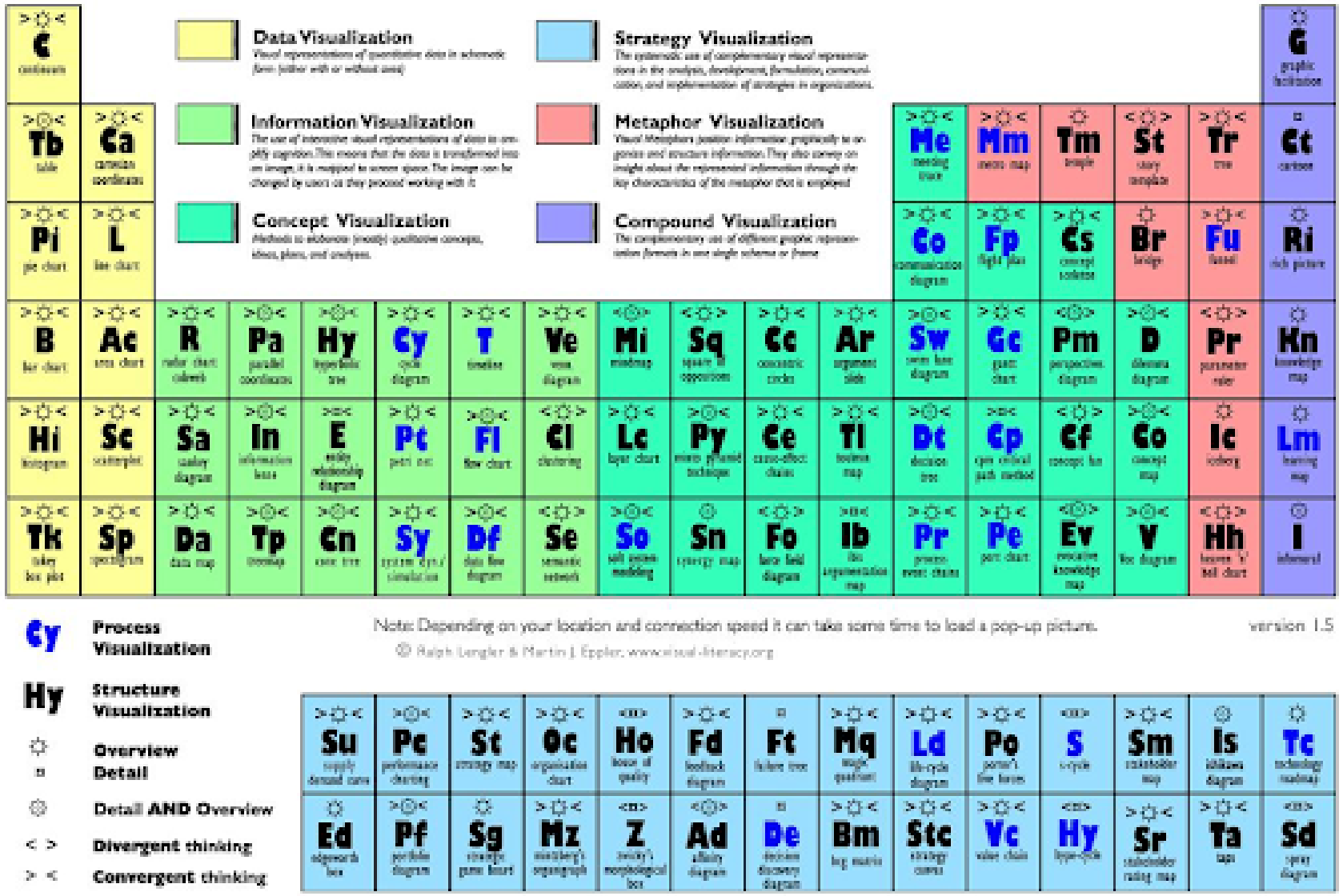
Efficient


Season Win Percent vs. Relative Payroll
Standard deviations above/below league average (15 team bins)
60%
55
50
45
40
Win percent
Oakland
-3
-2
-1
0
1
2
3
Standard deviations from average payroll
FIVETHIRTYEIGHT
BASED ON DATA FROM ESPN, BASEBALL PROSPECTUS, BASEBALL-REFERENCE

An amazing visual



A PERIODIC TABLE OF VISUALIZATION METHODS



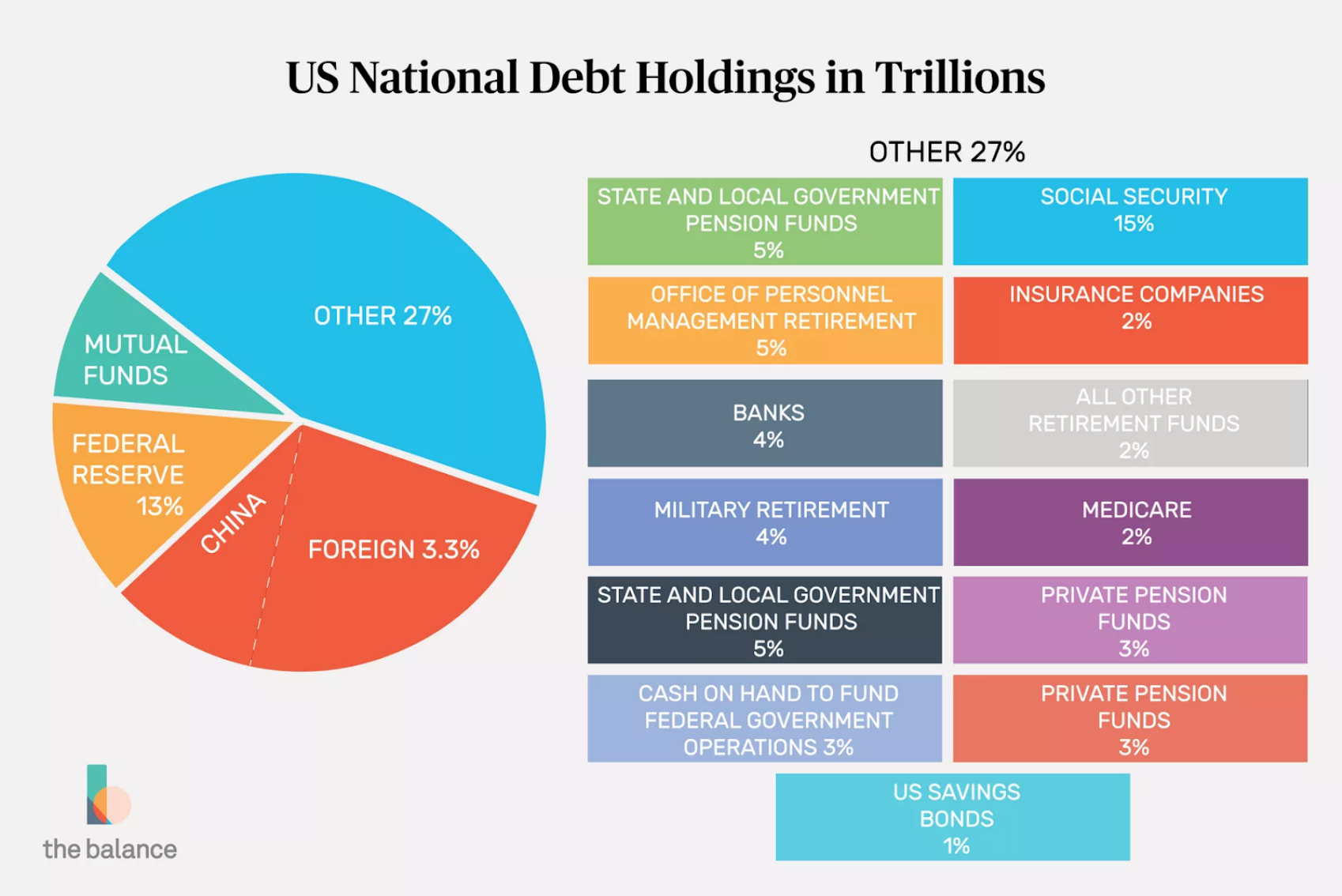
So terrible...

Maybe we do this instead...

Or this?

Or anything else!

Another atrocity



A template process

1. Write down your goal and intent for the visual
2. Gather the data that will help you achieve that goal
3. Decide how to tell your intended story with the data
4. Apply a visual representation of your data

Example

I worked on a project exploring the pay and education levels in different occupations in Nebraska relative to other states. I prepared the project by following the steps from the last slide:

1. Write down your goal and intent for the visual

My goal is to ***display wage and education patterns in occupations over time*** so that ***policymakers*** can ***understand the relationship between education and productivity in various job categories***.

Example

2. Gather the data that will help you achieve that goal

   Here is one of the SQL Statements that I used:

```
SELECT
    product/nobs AS wage,
    100.0*product/(
        SELECT SUM(product)
        FROM reducedOcc
        WHERE statefip=31
        GROUP BY year) AS percent
FROM reducedOcc
WHERE
    occ2010=1010
    AND statefip=31
GROUP BY year
ORDER BY year ASC
```

Example

3. Decide how to tell your intended story with the data
4. Apply a visual representation of your data

Let’s take a look at a webpage I made to explore and test my visuals:

http://dash4hank.herokuapp.com/

For Lab:

Using the data you extraced last week, work with your group to visualize the answers to your research question in Tableau. It will be useful to follow the steps described earlier for creating effective visualizations of data:

1. Write down your goal and intent for the visuals
2. Gather the data that will help you achieve that goal
3. Decide how to tell your intended story with the data
4. Create the visual representations of your data

Remember to focus on making visuals **Aesthetically Pleasing**, **Novel**, **Informative**, and **Efficient**