

A PRESENTATION FROM
DATA DOZEN
a visual analytics studio

ERIN WALDRON FROM DATA DOZEN · OCTOBER 4, 2022 · WIZARDS OF THE COAST · WORKSHOP #2

DEMYSTIFYING DATA

HOW WE MAKE DATA, WHEN TO USE IT, AND WHEN TO QUESTION IT



PARADOX

Data is super **simple** and super **complicated**.



DEMYSTIFYING DATA

Our intuition often reverses the complexity here.

DATA IS SIMPLE



DATA IS COMPLICATED



DEMYSTIFYING DATA

Our intuition often reverses the complexity here.
We assume the unfamiliar is more complicated.

DATA IS SIMPLE



DATA IS COMPLICATED



DEMYSTIFYING DATA

Our intuition often reverses the complexity here.
We assume the unfamiliar is more complicated.

TECHNOLOGY

DATA IS SIMPLE



CONTENT

DATA IS COMPLICATED



DEMYSTIFYING DATA

Our intuition often reverses the complexity here.
We assume the unfamiliar is more complicated.

TECHNOLOGY

DATA IS SIMPLE

Basic structure is just rows and columns.



CONTENT

DATA IS COMPLICATED



DEMYSTIFYING DATA

Our intuition often reverses the complexity here.
We assume the unfamiliar is more complicated.

TECHNOLOGY

DATA IS SIMPLE

Basic structure is just rows and columns.



CONTENT

DATA IS COMPLICATED

Because the world is complicated.



DEMYSTIFYING DATA

Our intuition often reverses the complexity here.
We assume the unfamiliar is more complicated.

TECHNOLOGY

DATA IS SIMPLE

Basic structure is just rows and columns.



CONTENT

DATA IS COMPLICATED

Because the world is complicated.



TIDY DATA

In his article “Tidy Data,” Hadley Wickham explains how we organize data so computers can read it.

ONE COLUMN PER VARIABLE

DIMENSIONS, MEASURES, FIELDS



ONE ROW PER OBSERVATION

RECORDS, TRANSACTIONS, EVENTS



ONE TABLE PER OBSERVATION TYPE

CONSISTENT GRANULARITY



TIDY DATA

Imagine a table in Excel.

TIDY DATA

One column per variable.

TIDY DATA

One column per variable.

DATE	TIME	AMOUNT	VENDOR

TIDY DATA

One row per observation.

DATE	TIME	AMOUNT	VENDOR

TIDY DATA

One row per observation.

DATE	TIME	AMOUNT	VENDOR
10/1/22	08:15 AM	\$8	Starbucks

TIDY DATA

One row per observation.

DATE	TIME	AMOUNT	VENDOR
10/1/22	08:15 AM	\$8	Starbucks
10/2/22	08:35 AM	\$8	Starbucks

TIDY DATA

One row per observation.

DATE	TIME	AMOUNT	VENDOR
10/1/22	08:15 AM	\$8	Starbucks
10/2/22	08:35 AM	\$8	Starbucks
10/2/22	08:42 AM	\$35	Gas Station

TIDY DATA

One row per observation.

DATE	TIME	AMOUNT	VENDOR
10/1/22	08:15 AM	\$8	Starbucks
10/2/22	08:35 AM	\$8	Starbucks
10/2/22	08:42 AM	\$35	Gas Station
10/2/22	07:08 PM	\$42	Dewey's Pizza

TIDY DATA

One row per observation.

DATE	TIME	AMOUNT	VENDOR
10/1/22	08:15 AM	\$8	Starbucks
10/2/22	08:35 AM	\$8	Starbucks
10/2/22	08:42 AM	\$35	Gas Station
10/2/22	07:08 PM	\$42	Dewey's Pizza
10/4/22	08:10 AM	\$8	Starbucks

TIDY DATA

One table per observational unit.

TABLE A

DATE	TIME	AMOUNT	VENDOR
10/1/22	08:15 AM	\$8	Starbucks
10/2/22	08:35 AM	\$8	Starbucks
10/2/22	08:42 AM	\$35	Gas Station
10/2/22	07:08 PM	\$42	Dewey's Pizza
10/4/22	08:10 AM	\$8	Starbucks

TABLE B

DATE	TIME	TO	FROM
10/2/22	08:12 AM	erin@datadozen.com	email
10/2/22	08:17 AM	erin@datadozen.com	email
10/2/22	08:26 AM	erin@datadozen.com	email

TIDY DATA

This basic rows and columns structure is underneath all our data visualizations.

ONE COLUMN PER VARIABLE

DIMENSIONS, MEASURES, FIELDS



ONE ROW PER OBSERVATION

RECORDS, TRANSACTIONS, EVENTS



ONE TABLE PER OBSERVATION TYPE

CONSISTENT GRANULARITY



THE HISTORY OF THE NFL

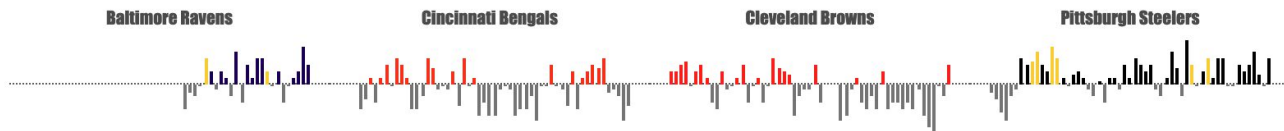
Seasons above or below .500 during the Super Bowl era 1966-2020. Gold represents Super Bowl win.

Hover over each bar to see detailed info from each season.

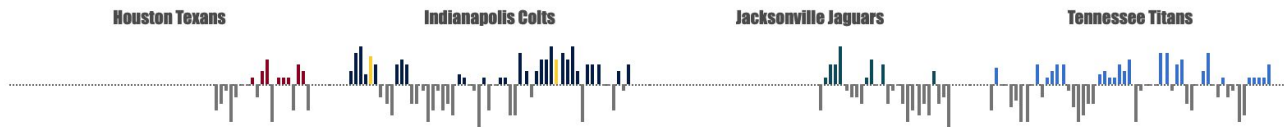
AFC East



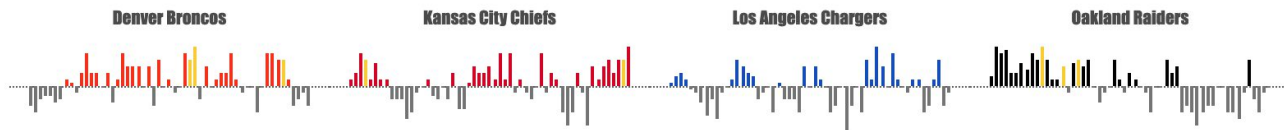
AFC North



AFC South



AFC West



SMALL MULTIPLES

THE HISTORY OF THE NFL

MATT CHAMBERS

2020

Full Data 1,610 rows 6 fields

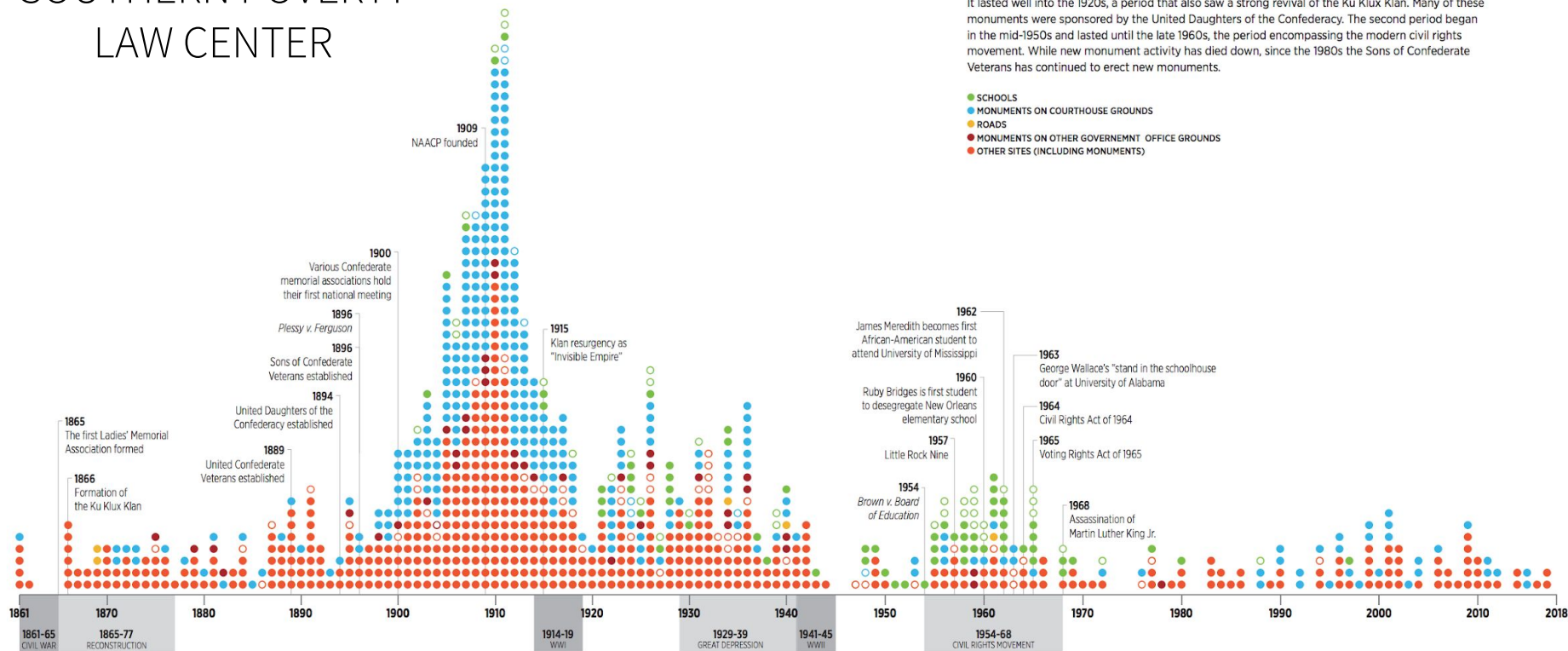
# Migrated Data Season	Abc Migrated Data Team	# Migrated Data Win	# Migrated Data Lose	# Migrated Data differential
1966	Arizona Cardinals	8	5	3
1967	Arizona Cardinals	6	7	-1
1968	Arizona Cardinals	9	4	5
1969	Arizona Cardinals	4	9	-5
1970	Arizona Cardinals	8	5	3
1971	Arizona Cardinals	4	9	-5
1972	Arizona Cardinals	4	9	-5
1973	Arizona Cardinals	4	9	-5
1974	Arizona Cardinals	10	4	6
1975	Arizona Cardinals	11	3	8

SMALL MULTIPLES

THE HISTORY OF THE NFL
MATT CHAMBERS
2020

FREQUENCY CHART

WHOSE HERITAGE SOUTHERN POVERTY LAW CENTER



FREQUENCY CHART

WHOSE HERITAGE SOUTHERN POVERTY LAW CENTER

Unique ID	feature_name	Honorees	City	County	State	Side	Coordinates	Symbol Type	Symbol Type Other	Sponsors
5146	Baruch Auditorium at VCU Medical College of Virginia	Simon Baruch	Richmond	Richmond County	VA		37.539375,-77.429460	Other	College	Unknown
5147	Dooley Hospital portico at VCU Medical College of Virginia	John Dooley	Richmond	Richmond County	VA		37.539169,-77.429765	Other	College	Unknown
5158	Wood Memorial Building at VCU Medical College of Virginia	Judson Wood	Richmond	Richmond County	VA		37.542271,-77.429562	Building		Unknown
5239	Lanier County Primary School	Sidney Lanier	Lakeland	Lanier	GA		31.038854205,-83.0755	School		Unknown
5238	Lake Lanier	Sidney Lanier	Lakeland	Lanier	GA		34.235667473,-83.9703	Other	Body of Water	Unknown
5236	Lanier High School	Sidney Lanier	Sugar Hill	Gwinnett	GA	Not Applicable	34.08580,-84.03320	School		Unknown
5237	Lanier County	Sidney Lanier	Lakeland	Lanier	GA		31.043213675,-83.0728	County/Municipality		Unknown
5260	Lanier Elementary School	Sidney Lanier	Tampa	Hillsborough	FL		27.883306555,-82.5231	School		Unknown
5242	Lanier County High School	Sidney Lanier	Lakeland	Lanier	GA		31.036121211,-83.0783	School		Unknown
5255	Sidney Lanier Center	Sidney Lanier	Gainesville	Alachua	FL		29.667832319,-82.3275	School		Unknown
5256	Sidney Lanier Bridge	Sidney Lanier	Brunswick	Glynn	GA		31.116887464,-81.4847	Other	Bridge	Unknown
5241	Lanier County Middle School	Sidney Lanier	Lakeland	Lanier	GA		31.037661301,-83.0782	School		Unknown
5240	Lanier County Elementary School	Sidney Lanier	Lakeland	Lanier	GA		31.034566294,-83.0742	School		Unknown
5512	Lanier Middle School	Sidney Lanier	Buford	Gwinnett	GA	Not Applicable	34.14796,-84.06533	School		
5513	Lanier Elementary	Sidney Lanier	Tulsa	Tulsa	OK	Not Applicable	36.13585,-95.93986	School		
5514	Lanier High School	Sidney Lanier	San Antonio	Bexar	TX	Not Applicable	29.42219,-98.51538	School		
5432	Lanier Ct.	Sidney Lanier	Montgomery	Montgomery	AL	Not Applicable	32.354767,-86.313247	Roadway		
5624	Lanier County School District	Sidney Lanier	Lakeland	Lanier	GA	Not Applicable	31.03974,-83.07466	Other	School District	
5771	Leconte Hall at University of Georgia	Joseph LeConte	Athens	Clarke	GA	Not Applicable	33.95338,-83.37334	Building		
5187	Lane Street	James Henry Lane	New Orleans	Orleans	LA		30.008924691,-90.1077	Roadway		Unknown
5189	Walker Street	John George Walker	New Orleans	Orleans	LA		30.017905275,-90.1063	Roadway		Unknown

TIDY DATA

At the end of the day, it's all just columns and rows.

DEMYSTIFYING DATA

So where do things get complicated?

TECHNOLOGY

DATA IS SIMPLE

Basic structure is just rows and columns.



CONTENT

DATA IS COMPLICATED

Because the world is complicated.



DEMYSTIFYING DATA

So where do things get complicated?

TECHNOLOGY

DATA IS SIMPLE

Basic structure is just rows and columns.



CONTENT

DATA IS COMPLICATED

Because the world is complicated.



LISA GITELMAN

Raw data is an oxymoron.



OUR HUMAN FILTER

We **create** every measurement in our dataset.



CONFIDENCE

You are fully equipped to ask questions about a dataset.



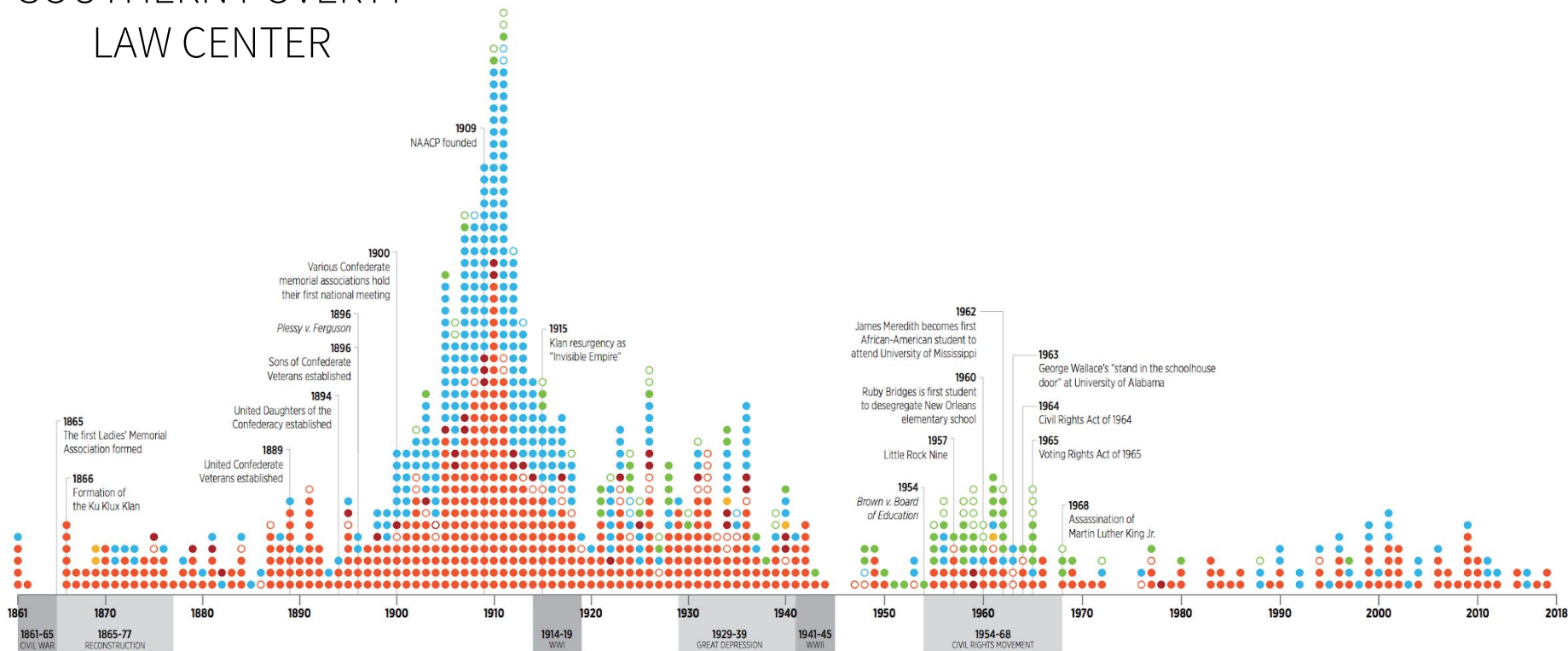
FREQUENCY CHART

WHOSE HERITAGE SOUTHERN POVERTY LAW CENTER

DEFINITIONS

WHAT'S A MONUMENT?

And who gets to decide that definition?



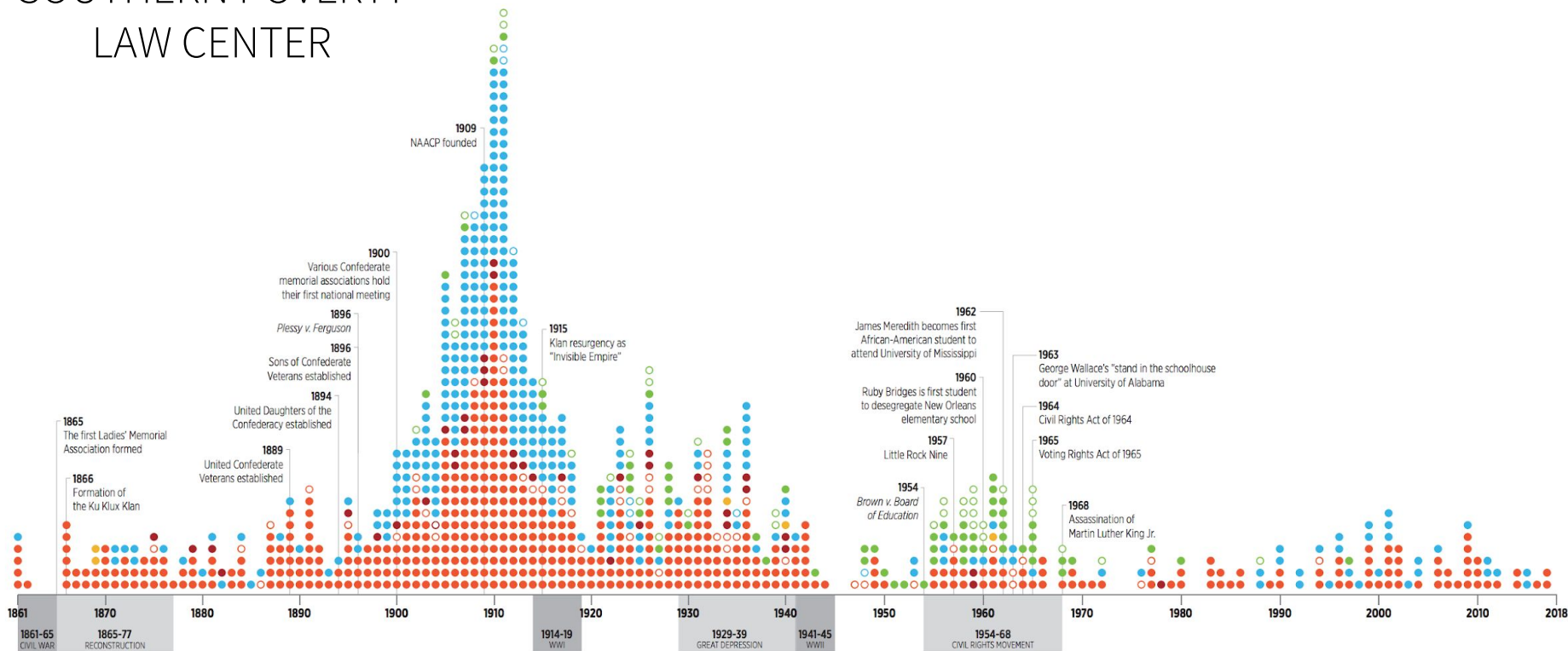
FREQUENCY CHART

WHOSE HERITAGE SOUTHERN POVERTY LAW CENTER

SAMPLES

WHAT DO WE INCLUDE?

Therefore, what (or who) do we exclude?



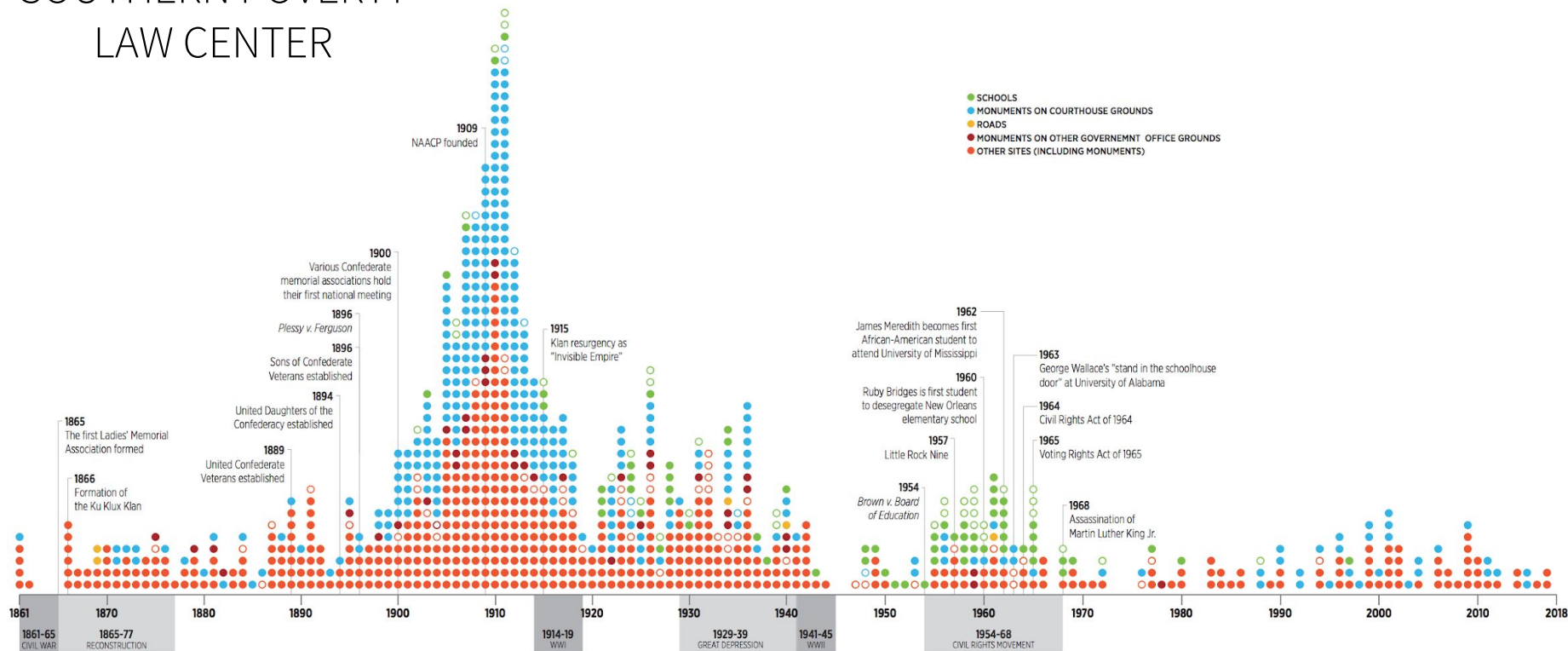
FREQUENCY CHART

WHOSE HERITAGE SOUTHERN POVERTY LAW CENTER

CATEGORIES

HOW TO GROUP THINGS?

How do our categories restrict the answers.



EUROPE

SHARE

The Scale of the Migrant Crisis, From 160 to Millions

By GREGOR AISCH, SARAH ALMUKHTAR, JOSH KELLER and WILSON ANDREWS UPDATED SEPT. 22, 2015



4

ARTICLES REMAINING

AREA CHART

THE SCALE OF THE MIGRANT CRISIS

NEW YORK TIMES

2015

METHODOLOGY

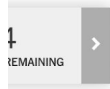
HOW DO WE COUNT?

And how good are our estimates?

This photo shows about **160** migrants in southern Hungary waiting to board a bus to a registration center on Sept. 9. They will be excluded from the only two relocation plans approved by European leaders so far. The plans will relocate a total of **160,000** people in Greece and Italy to other countries in Europe.

160
migrants in
the photo

160,000
migrants
relocated in
approved plans



The plans add up to the largest agreement ever to distribute asylum seekers among member states. But they address only a small fraction of the expected total.

AREA CHART

THE SCALE OF THE MIGRANT CRISIS

NEW YORK TIMES

2015

160
migrants in
the photo

160,000
migrants
relocated in
approved plans

549,000
migrant arrivals
in Greece, Italy
and Hungary

1,300,000
asylum
applications
expected this
year

4,700,000
asylum seekers in
Turkey, Lebanon and
Jordan

POWER & BIAS

AND WHO'S COUNTING?

Impacts of the perspective of the counter.

AREA CHART

THE SCALE OF THE MIGRANT CRISIS

NEW YORK TIMES

2015

Squares show the relative sizes of the numbers and are not meant to convey that they are parts of a whole.

THE AGE of BORDERS

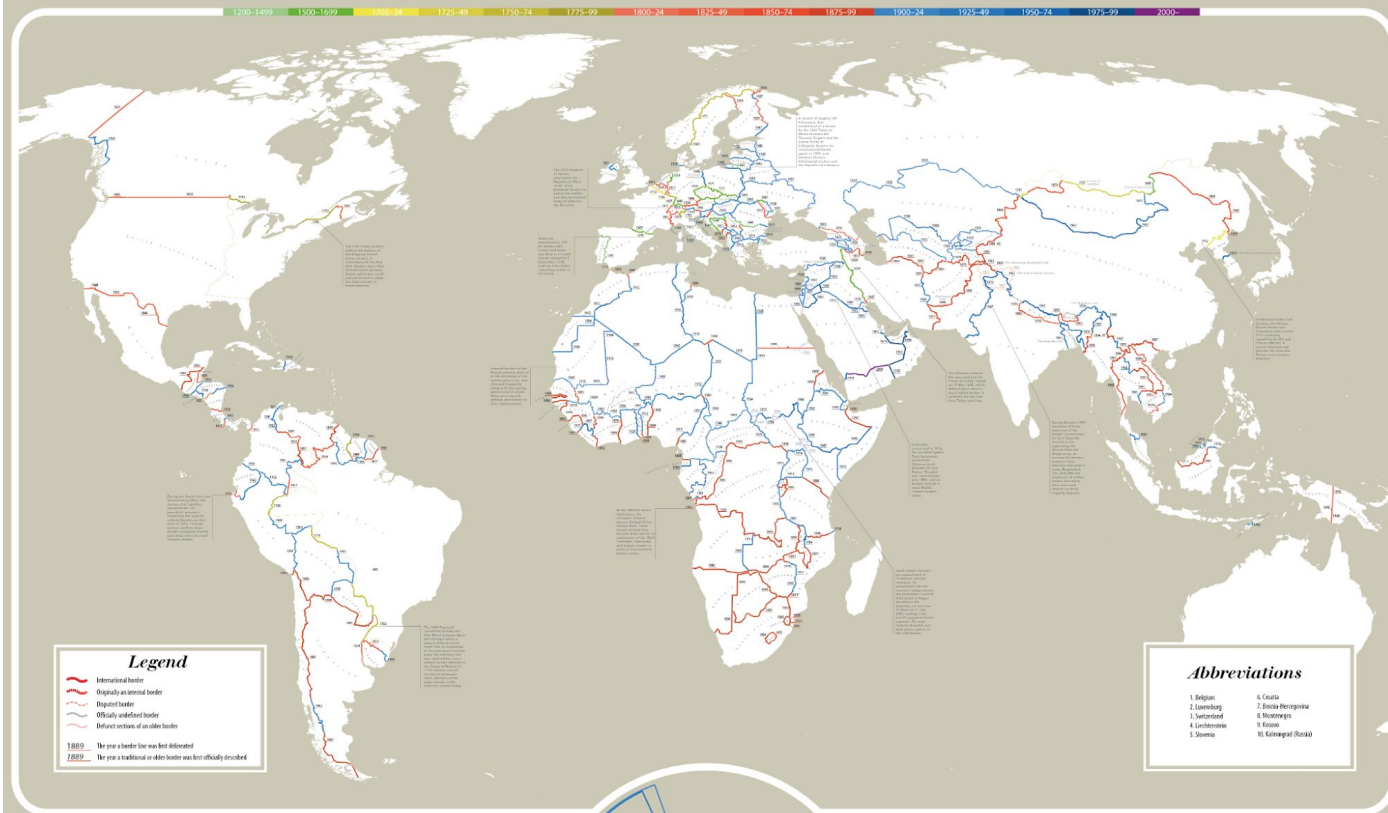
When were the world's current borders first defined?

MAP

THE AGE OF BORDERS

PISSEGURI82

2018



Average border age by region:

Borders younger than 100 years (33,6%):

AMBIGUITY

WHO'S RIGHT?

Data that depends on who you ask.

MAP

THE AGE OF BORDERS

PISSEGURI82

2018

Legend



International border



Originally an internal border



Disputed border



Officially undefined border



Defunct sections of an older border

1889

The year a border line was first delineated

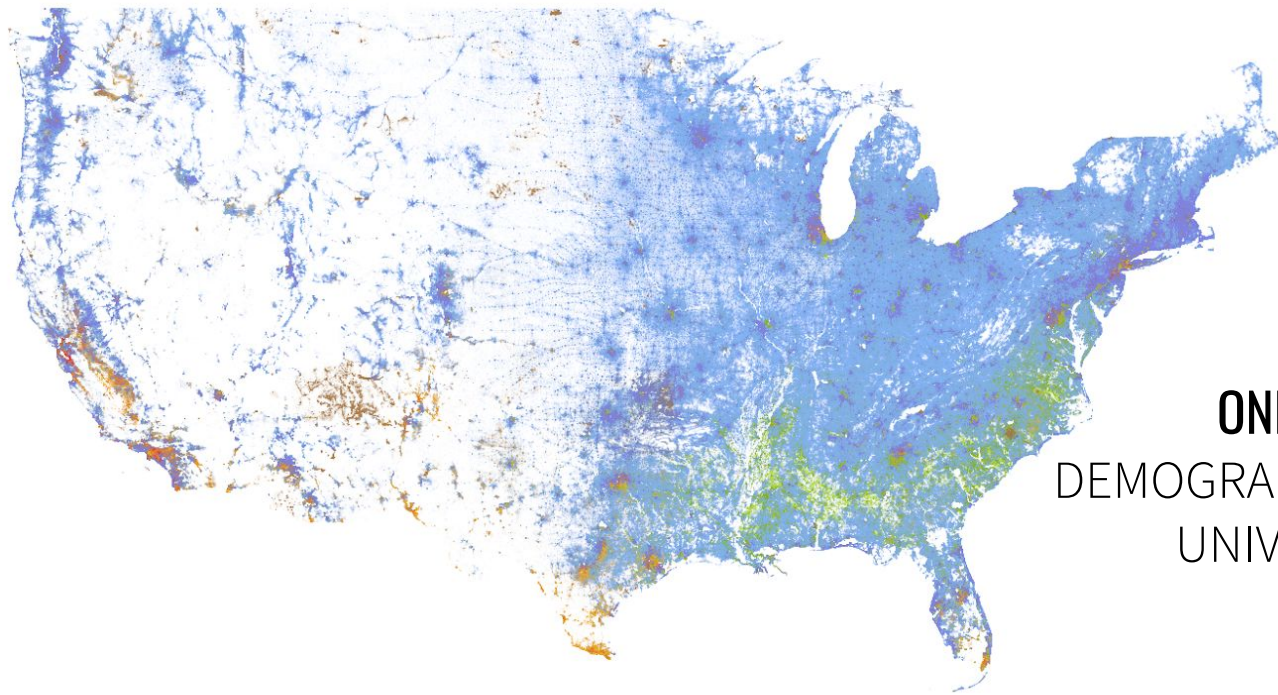
1889

The year a traditional or older border was first officially described

DID WE RECORD THAT?

There's a difference between 0 and null.





MAP

ONE PERSON, ONE DOT

DEMOGRAPHICS RESEARCH GROUP

UNIVERSITY OF VIRGINIA

2010

IS THIS REPRESENTATIVE?

How much complexity must be retained?



WHY WE HAVE REMOVED THE RACIAL DOT MAP

After nine years and millions of views, the 2010 Racial Dot Map has reached its expiration date. We have taken it offline as it no longer provides the most accurate depiction of the nation's population distribution and changing racial composition. Several factors contributed to this decision:

1. The 2020 Census count released by the U.S. Census Bureau on August 12, 2021 provides a new snapshot of the U.S. population by race and ethnicity, making 2020 the most current data of record. As demographers committed to data integrity, we cannot continue to host a map that does not accurately tell the story of race in the United States.
2. Between the 2010 and 2020 censuses, the multi-race and "Some Other Race" categories show significant growth. In the 2020 Census, ten percent of the population identified as multi-race compared to three percent in 2010; and "Some Other Race" became the second largest racial group, surpassing the population identifying themselves as Blacks or African-Americans. Both the dynamic growth of these populations and complexity of reflecting this rich diversity through color-coded dots made the model used for the Racial Dot Map inadequate to the task.
3. Producing a new map, equally elegant in its simplicity but capable of reflecting many more racial/ethnic groups is beyond our organization's financial and personnel resources.

We appreciate that so many of you have been passionate advocates for the 2010 Racial Dot Map and the ways it has helped to promote equity in your communities.

THE HISTORY OF THE NFL

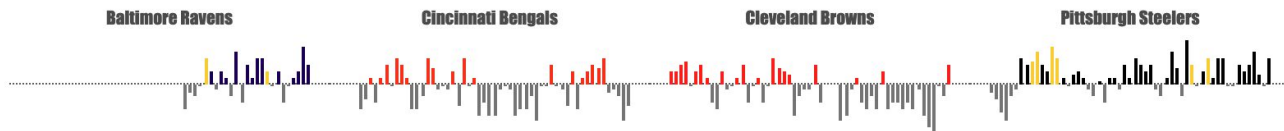
Seasons above or below .500 during the Super Bowl era 1966-2020. Gold represents Super Bowl win.

Hover over each bar to see detailed info from each season.

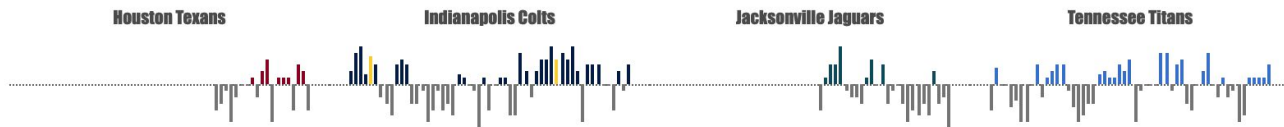
AFC East



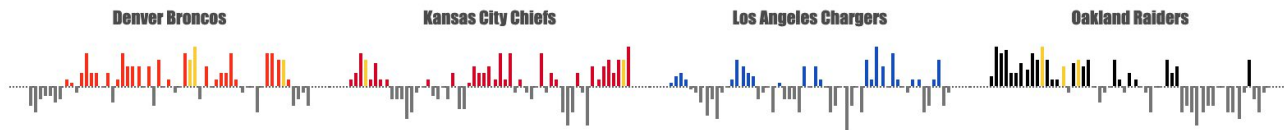
AFC North



AFC South



AFC West



SMALL MULTIPLES

THE HISTORY OF THE NFL

MATT CHAMBERS

2020

CONSISTENCY

HOW DO THINGS CHANGE?

How do we account of differences over time?

Team history

- Carolina Panthers (1995–present)

NFC South

Atlanta Falcons

Carolina Panthers

New Orleans Saints

Tampa Bay Buccaneers

NFC West

Arizona Cardinals

Los Angeles Rams

San Francisco 49ers

Seattle Seahawks

SMALL MULTIPLES

THE HISTORY OF THE NFL

MATT CHAMBERS

2020

Team history

Since 1920:^[5]

- Racine Cardinals (1920–1921)
- Chicago Cardinals (1922–1943, 1945–1959)
- Card-Pitt (1944)
- St. Louis Cardinals (1960–1987)
- Phoenix Cardinals (1988–1993)
- Arizona Cardinals (1994–present)

1921, existed prior to the formation of the NFL. The person keeping the minutes of the first league meeting, unfamiliar with the nuances of Chicago football, recorded the Cardinals as from **Racine**, **Wisconsin**. The team was renamed the "Chicago Cardinals" in 1922 after a team actually from **Racine**, Wisconsin (the **Horlick-Racine Legion**) entered the league. That season the team moved to **Comiskey Park**.

Full Data 1,610 rows 6 fields

# Migrated Data Season	Abc Migrated Data Team	# Migrated Data Win	# Migrated Data Lose	# Migrated Data differential
1966	Arizona Cardinals	8	5	3
1967	Arizona Cardinals	6	7	-1
1968	Arizona Cardinals	9	4	5
1969	Arizona Cardinals	4	9	-5
1970	Arizona Cardinals	8	5	3
1971	Arizona Cardinals	4	9	-5
1972	Arizona Cardinals	4	9	-5
1973	Arizona Cardinals	4	9	-5
1974	Arizona Cardinals	10	4	6
1975	Arizona Cardinals	11	3	8

STANDARDIZATION

WHAT WAS ERASED?

What inconsistencies are simply deleted?

SMALL MULTIPLES

THE HISTORY OF THE NFL

MATT CHAMBERS
2020

THE HISTORY OF THE NFL

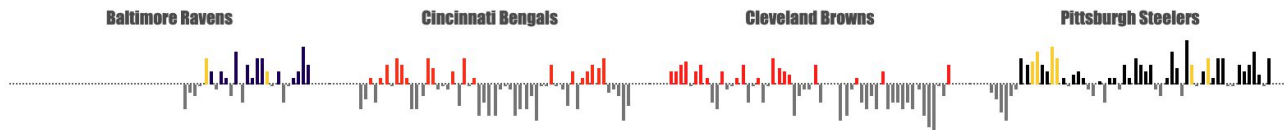
Seasons above or below .500 during the Super Bowl era 1966-2020. Gold represents Super Bowl win.

Hover over each bar to see detailed info from each season.

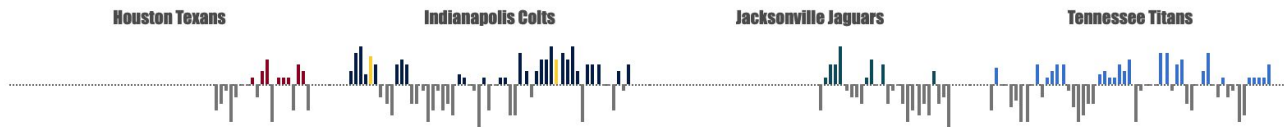
AFC East



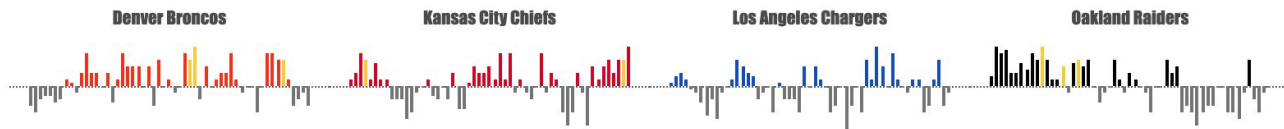
AFC North



AFC South



AFC West



PRESENTISM

WHAT'S OUR LENS TODAY?

Our work is always of our time.

SMALL MULTIPLES

THE HISTORY OF THE NFL

MATT CHAMBERS

2020

HOW WE TALK ABOUT DATA

WHAT WE SAY

Our language implies an inaccurate level of objectivity.

- ▶ “Raw data”
- ▶ “Hard data”
- ▶ “Data-driven decisions”
- ▶ “The data doesn’t lie”

WHAT WE MEAN

Data can contribute to our ways of knowing.

- ▶ original data / underlying data
- ▶ quantitative information
- ▶ less subjective / more informed decisions
- ▶ data can see things humans can’t

WHY WE NEED DATA

WHAT WE SEE

Humans experience the world from a fixed, singular viewpoint.

- ▶ one place, one time
- ▶ limited and fallible memory
- ▶ cognitive biases
- ▶ individual experience

WHAT THE DATA SEES

Data provides information from non-human viewpoints.

- ▶ broader patterns, individual outliers
- ▶ (more) consistent record
- ▶ (typically) less biases
- ▶ (more) common and collective form of truth

ACTIVITY #2 | DATA MAKER

WHAT YOU NEED

- ▶ Game

This could be any game you love to play. (Personally, I think Magic cards work fantastic for this exercise!)

- ▶ Spreadsheet

This could be a google sheet or excel file, or just a big piece of paper where you draw some lines.

WHAT TO DO

- ▶ 1) Pick your content

Pick an aspect of your favorite game to capture as a dataset. (Individual game pieces, or individual turns of people playing the game, etc.)

- ▶ 2) Make columns and rows

Make one column for each variable you want to record, and populate at least 10 rows of data.

- ▶ 3) Jot down the hard parts

Take note of the difficult decisions or judgement calls you're making as you create your dataset.

ACTIVITY #2 | DATA MAKER

Here's a starter example with Magic cards.

NAME	TYPE	COLOR	ATTACK	DEFEND
Valorous Steed	Creature	White	3	3
Clockwork Drawbridge	Artifact Creature	White	0	3
Plains	Basic Land	White		
Shambling Goblin	Creature	Black	1	1
Release the Dogs	Sorcery	White		

ACTIVITY #2 | DATA MAKER

Here's a starter example with Scrabble.

PLAY	PLAYER	WORD	VALUE
1	Erin	dwel	18
2	Jay	uptown	45
3	Erin	jeers	12
4	Jay	quin	24
5	Erin	grout	12

DAY 1



DATAVIZ 101

what it is, what it's
not, and what you
can do with data
visualization



DEMYSTIFYING DATA

how we make data,
when to use it
and when to
question it

DAY 2



DESIGN DECISIONS

the functional
impact of artistic
design in audience
engagement



COMMUNICATION CHECKLIST

how dataviz can
make us better
communicators and
collaborators

HERE'S WHAT WE'LL DO TOGETHER

ERIN WALDRON FROM DATA DOZEN · CONTACT@DATADOZEN.COM · 913-303-1125

THANK YOU!

WWW.DATADOZEN.COM

