

MAKING PEOPLE CARE

PAIRING NARRATIVE WITH VISUALIZATION

ILIA BLINDERMAN

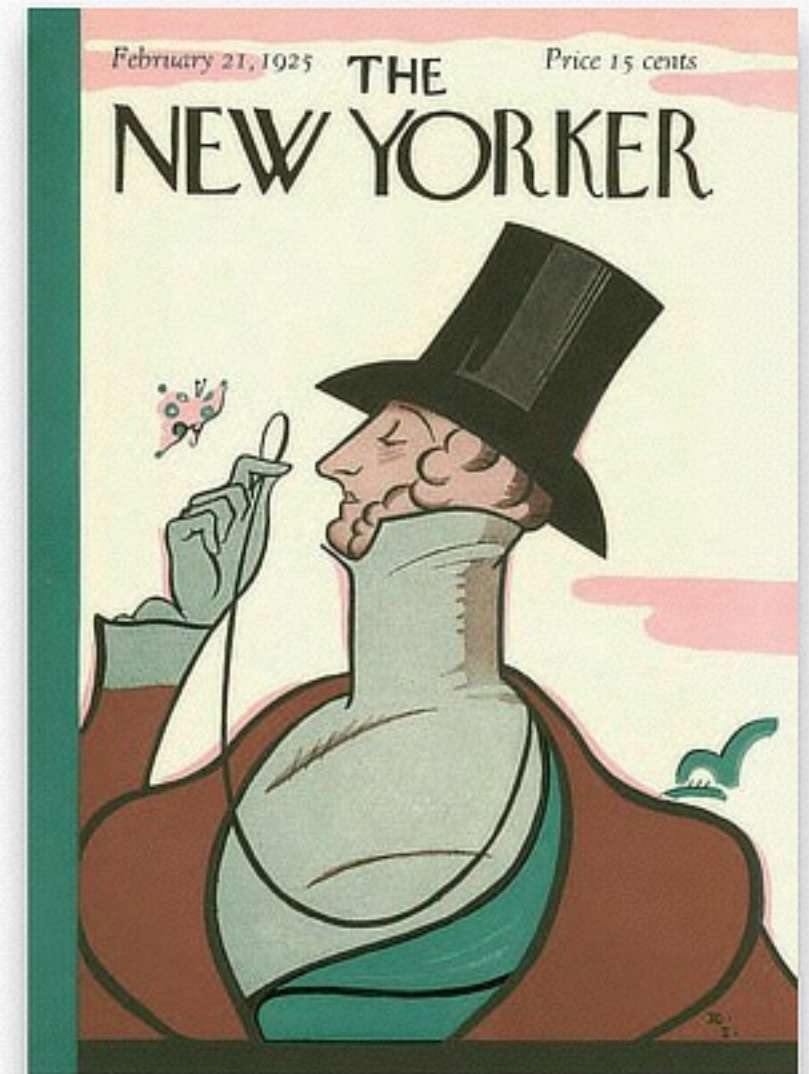
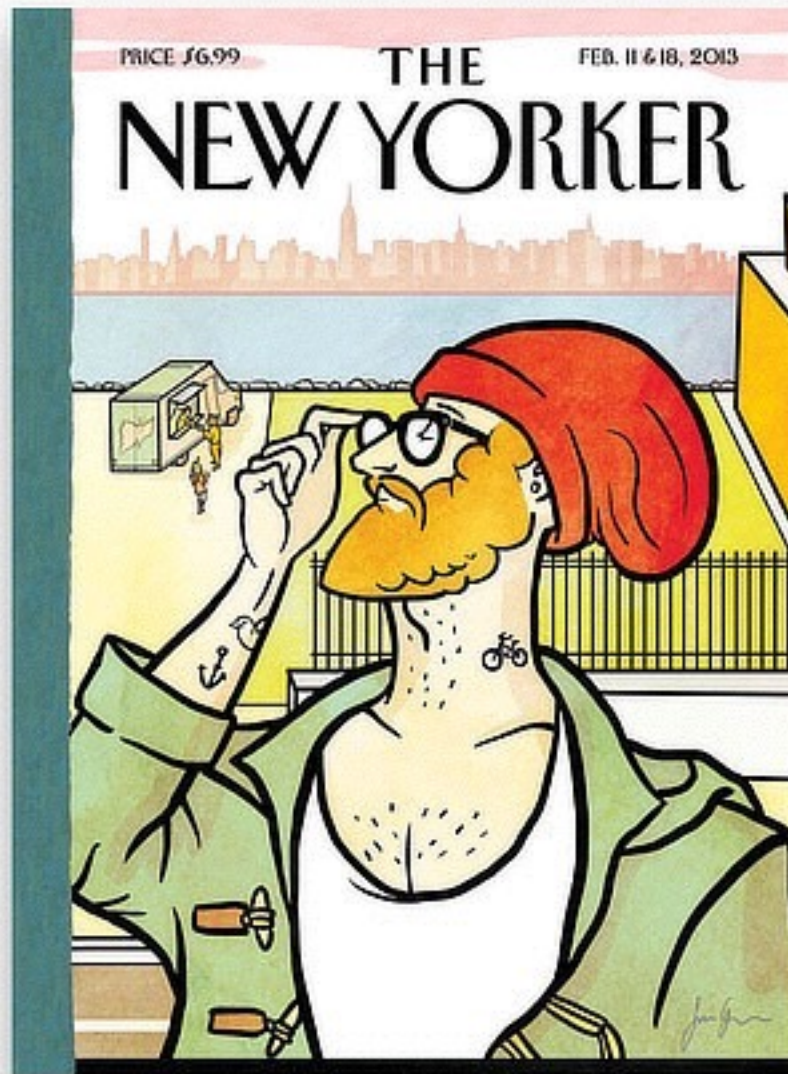
STRUCTURE

- ▶ My background
- ▶ Some theory
- ▶ Outline of missed connections project + code examples
- ▶ Discussion of pairing narrative

MY BACKGROUND

- ▶ Psychological research
- ▶ Journalism
- ▶ Data + viz

THE NEW YORKER MAGAZINE



EXAMPLES OF NEW YORKER NARRATIVES

- ▶ [Structure of long form writing](#) - 7,597 words (38 min.)
- ▶ [Commuting](#) - 6,705 words (34 min.)
- ▶ [Elevators](#) - 8,100 words (41 min.)

WHY READ ABOUT ELEVATORS FOR AN HOUR?

- ▶ Inherent interest: Because you're all about elevators
- ▶ Emotional resonance: Because the story garners some emotional response
- ▶ Which of the two is the superior strategy?

TEXT

ELEVATORS?



THE WAY YOU COMMUNICATE FACTS MATTERS

- ▶ Communicating > presenting facts
- ▶ 1) Capture audience attention
- ▶ 2) Retain it.

WAIT, WHAT ABOUT THESE MISSED CONNECTIONS?

CL

new york

>

all new york

>

personals

>

missed connections

[account] post

missed connections

<< search missed connections

Q

save search

☐ search titles only

☐ has image

☐ posted today

☐ include nearby areas

MILES FROM ZIP

miles

from zip

AGE

min

max

reset

update search

list

thumb

gallery

map

<<

< prev

1 to 100 of 759

next >

newest

★

Mar 27

Cute guy next to me during BvS movie tonight - m4m (City Center 15 Cinema)

map

✕

★

Mar 27

FLESHLIGHT GIRLS ENTIRE COLLECTION FLESHLIGHT - m4w pic map

✕

★

Mar 27

unthinkably cruel

✕

★

Mar 27

You just got off the 6 at Bleecker maybe 20-25 minutes ago - w4m (Greenwich Village)

map

✕

★

Mar 27

Looking for the NYFC fan with the lovely smile - w4m 27 (6 train)

map

✕

★

Mar 27

FLESHLIGHT GIRLS ENTIRE COLLECTION FLESHLIGHT - m4w pic map

✕

★

Mar 27

Asian girl on the F train at Delancey St - m4w (Lower East Side)

✕

★

Mar 27

Re: The Krusty Krab - m4w pic

✕

★

Mar 27

To the most beautiful woman with the beautifully flowered name - m4w

✕

★

Mar 27

ANR/ABF - w4m 35 (Brooklyn)

✕

★

Mar 27

train/dream boy - w4m map

✕

★

Mar 27

My Panty Sniffer - w4m

✕

★

Mar 27

K Pop in Midtown - m4w (Midtown)

map

✕

WHY GO DEEP ON MISSED CONNECTIONS?

- ▶ Interesting
- ▶ No one had done it
- ▶ Rich dataset
- ▶ Free (ish) from observer effects

HOW?

- ▶ Getting data: Python, BeautifulSoup
- ▶ Cleaning + analyzing data: R; Python's Natural Language Toolkit (NLTK) Package
- ▶ Visualizing data: D3.js

COLLECTING DATA

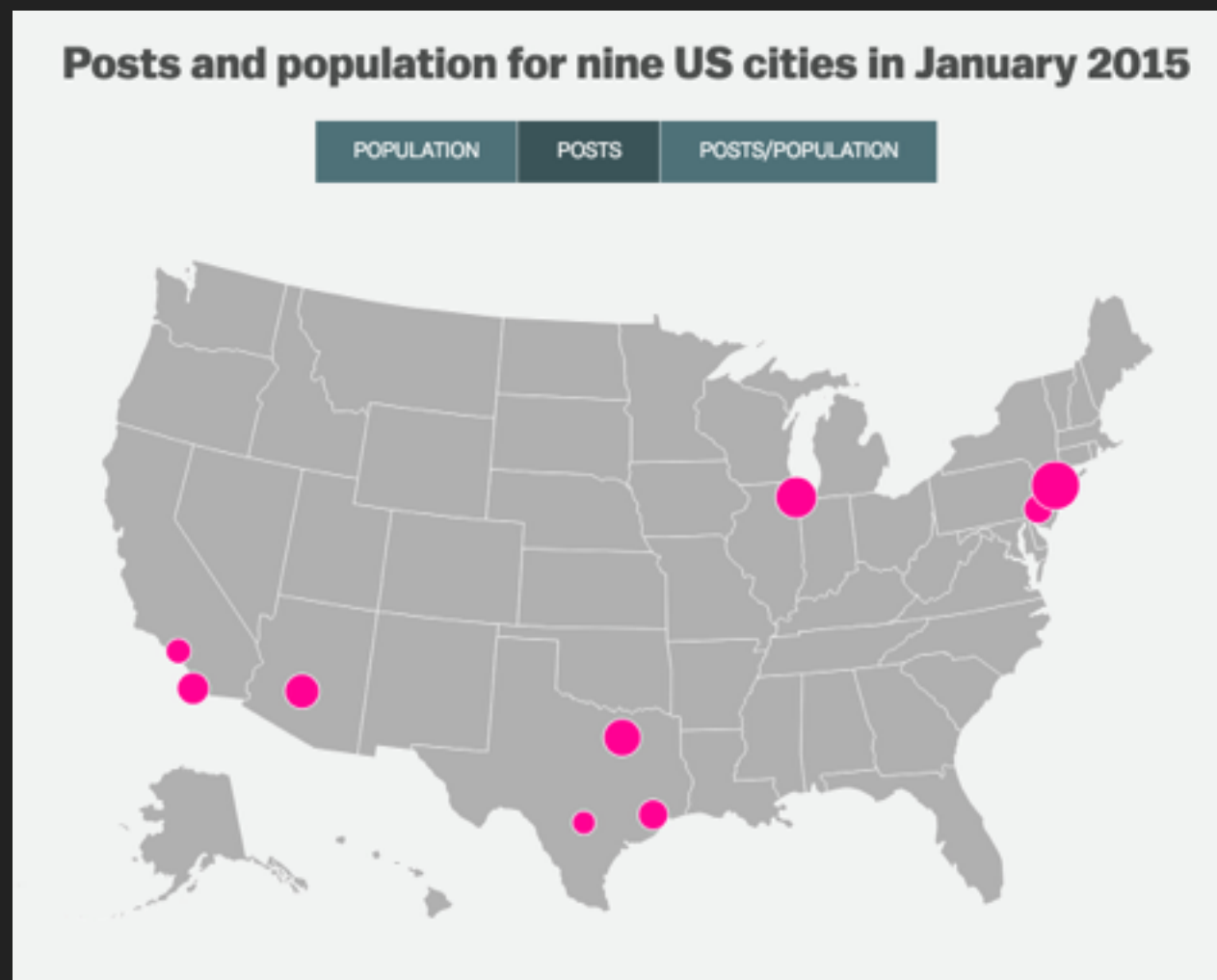
- ▶ [Python script](#) (c/o [Brian Abelson](#))
- ▶ Daily manual scraping during January 2015
- ▶ ~12,000 missed connections records

DATA POINTS

- ▶ Who was looking for whom
- ▶ Age of poster
- ▶ Time of posting
- ▶ City
- ▶ Date
- ▶ Length + contents of message

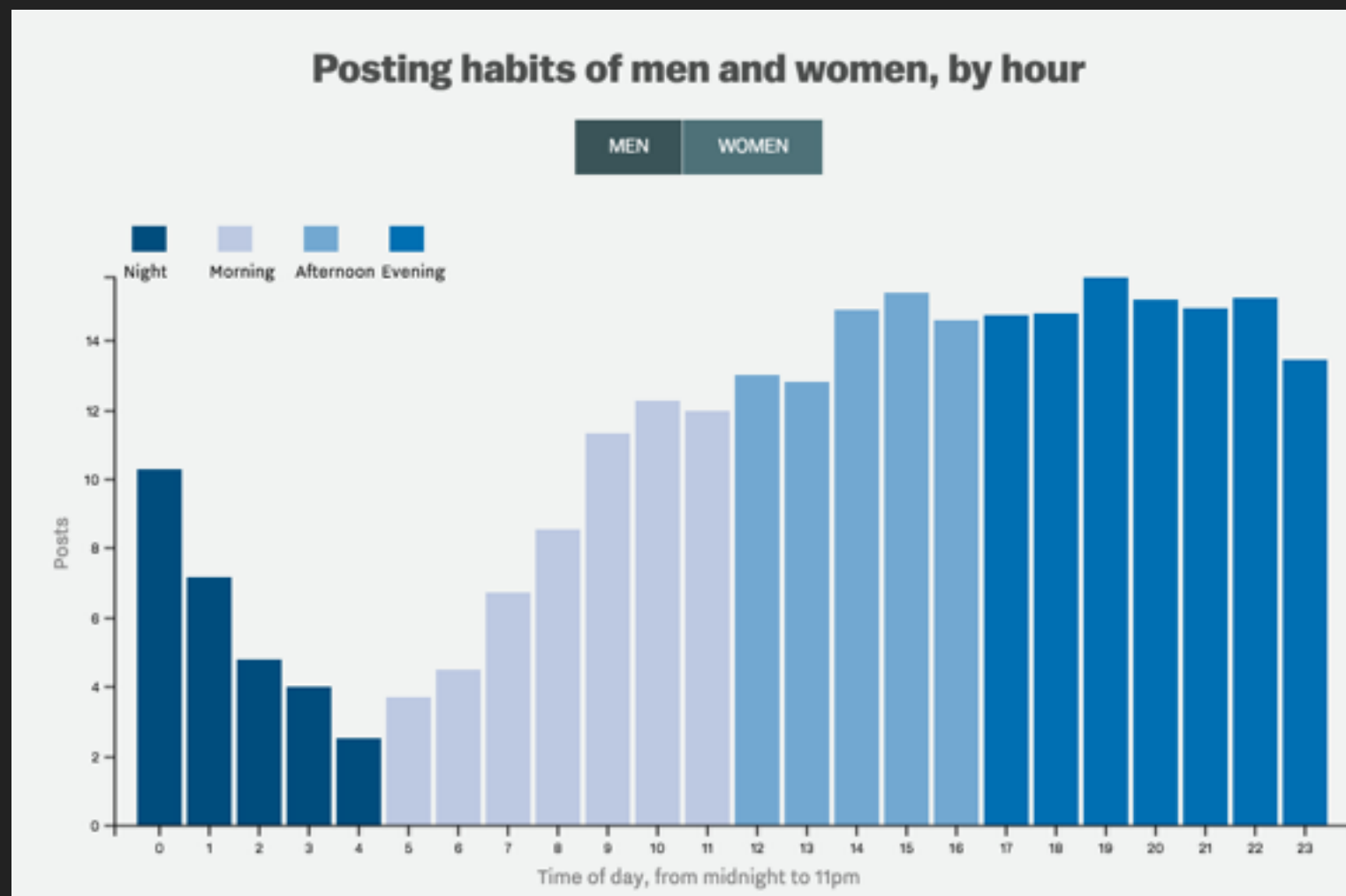
FINDINGS

- ▶ NYC has the most posts, but Dallas has the highest posts per capita



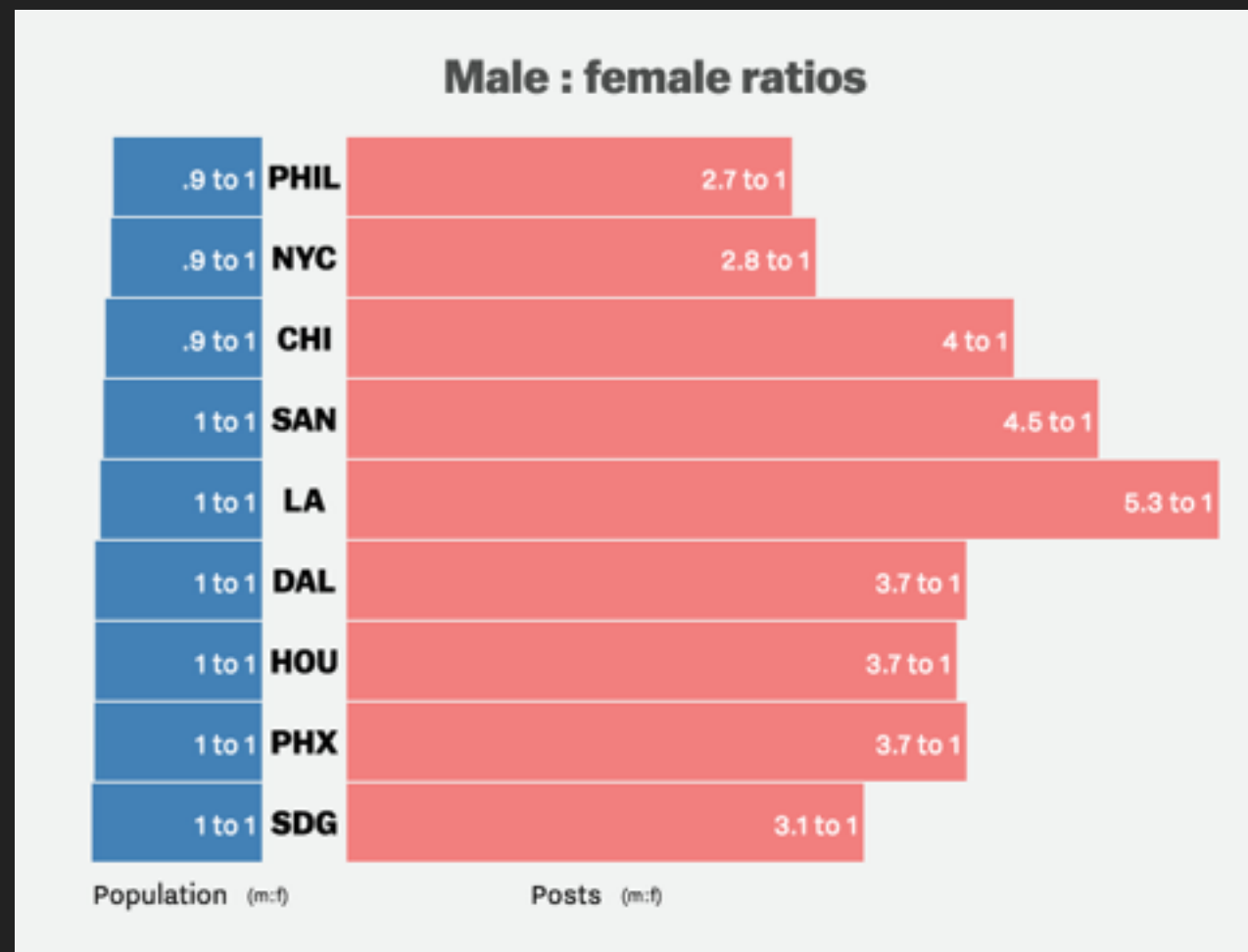
FINDINGS

- ▶ Men post all day, women post after work



FINDINGS

- ▶ Men post much, much, much more than women.



FINDINGS

- ▶ Everyone uses the term “eye contact”



VIZ – INDIVIDUAL CHARTS

- ▶ How to visualize each part of the story?
 - ▶ This is hard.
 - ▶ Even Mike Bostock, the creator of D3, wrestles with this.
- ➡ ▶ <http://www.datavizcatalogue.com/>
- ➡ ▶ <http://bl.ocksplorer.org/>
- ➡ ▶ <http://christopheviau.com/d3list/>

EXAMPLE

► How to visualize pervasive differences in

1. Age
2. Individual post length
3. Number of posts
4. Missed connection poster sex
5. Missed connection subject sex
6. City

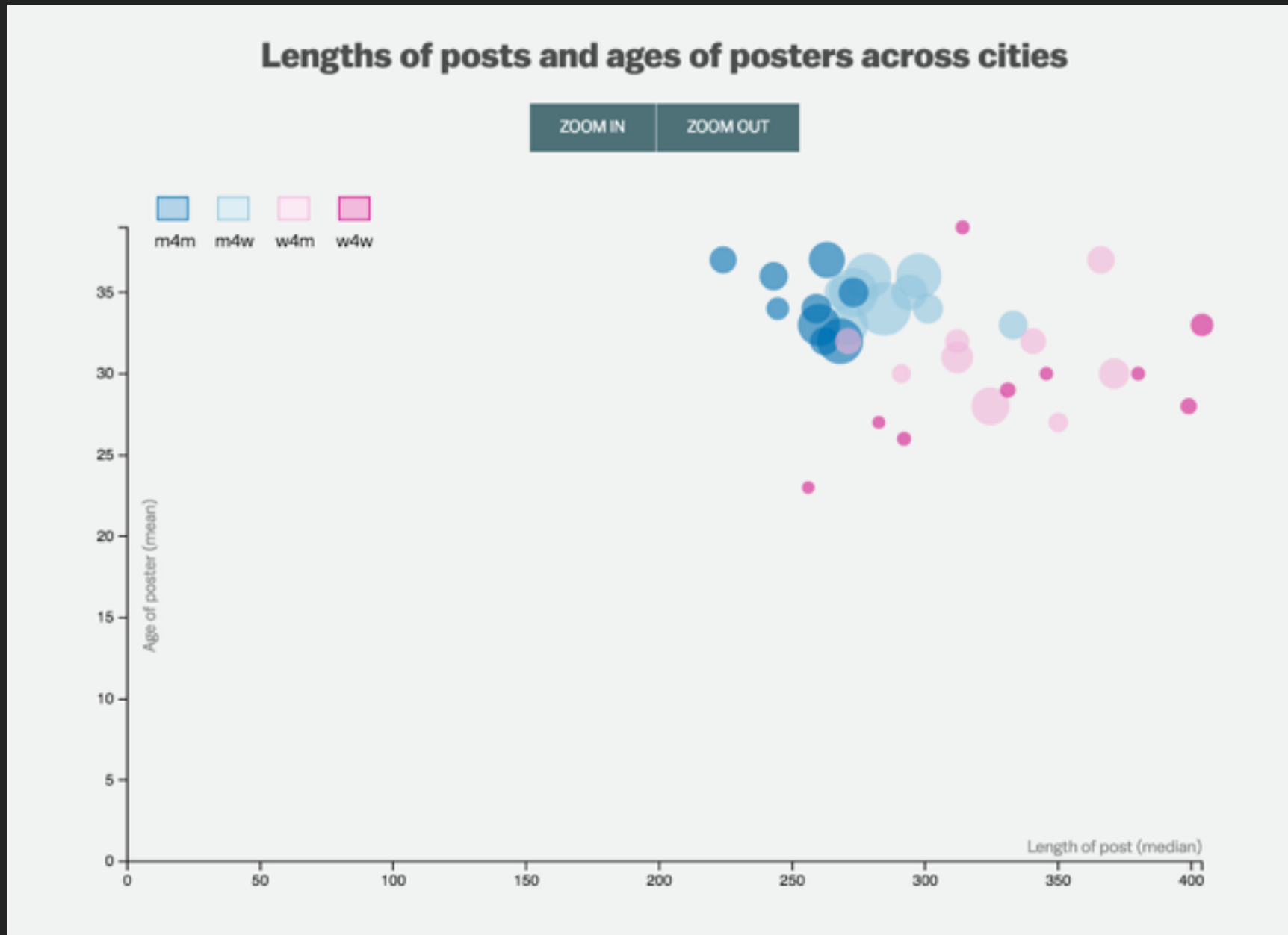
UHHHH



SCATTERPLOTS ARE AMAZING – LOTS OF VISUAL CHANNELS

- ▶ X axis
- ▶ Y axis
- ▶ Circle radius
- ▶ Circle color
- ▶ Interactive aspects

SCATTERPLOT



MAKING CIRCLES

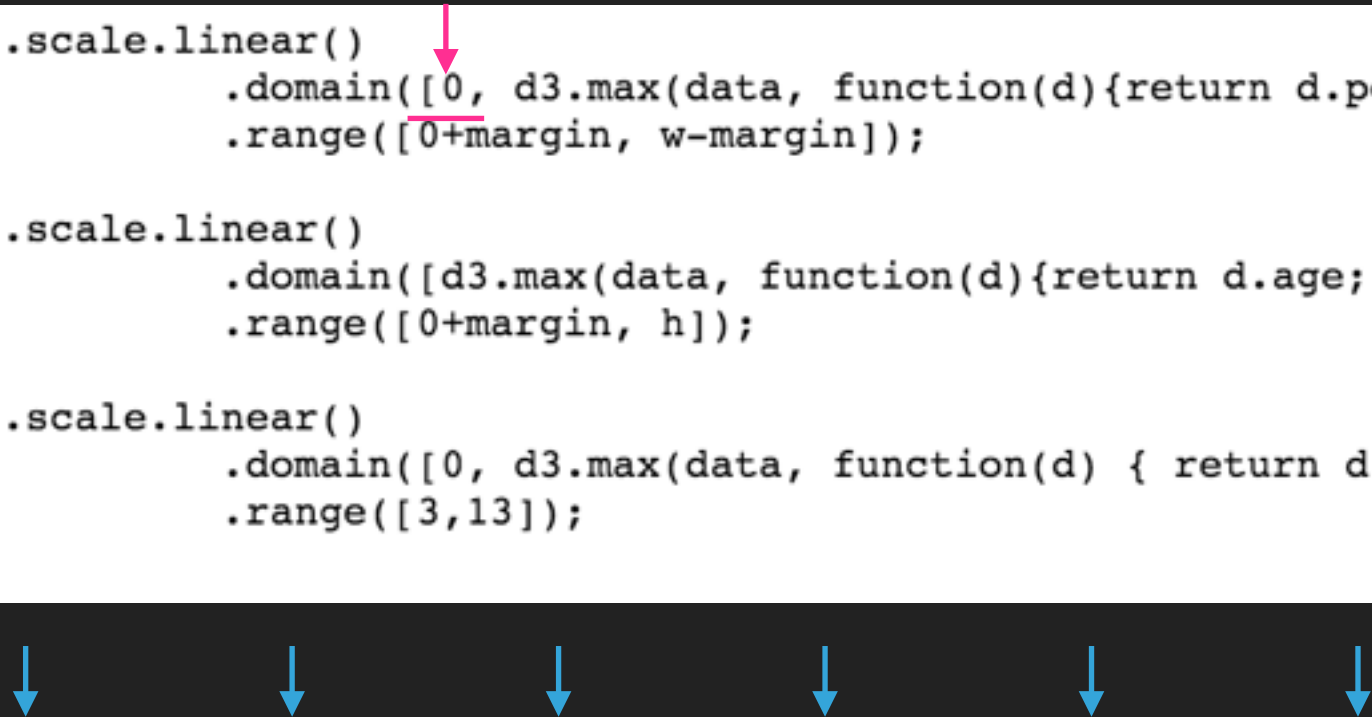
```
var circles = svg.selectAll("circle")
    .data(data)
    .enter()
    .append("circle")
    .attr({
        cx: function(d) {return xScale(d.post_length);},
        cy: function(d) {return yScale(d.age);},
        r: function(d){return rScale(d.count);},
        class: function(d) {return d.city.replace(/\s/g, ' ')},
        id: function(d) {return d.seeker}
    })
    .style({
        fill: function(d) {return colScale(d.seeker);},
        opacity: 0.6
    });
```


SCATTERPLOT DATA – JOINED TO VISUAL ELEMENTS

	city	seeker	Median_body_length	mean_age	median_topic_length	count
1	Chicago	m4m	260	33	25	579
2	Chicago	m4w	273	35	22	800
3	Chicago	w4m	312	31	22	311
4	Chicago	w4w	331	29	23	36
5	Dallas	m4m	263	37	22	415
6	Dallas	m4w	278.5	36	22	693
7	Dallas	w4m	371	30	19	279
8	Dallas	w4w	345.5	30	23	15
9	Houston	m4m	273	35	22	265
10	Houston	m4w	294	35	23	410
11	Houston	w4m	312	32	22	158
12	Houston	w4w	292	26	19	22
13	Los Angeles	m4m	262	32	26	236
14	Los Angeles	m4w	333	33	29	244
15	Los Angeles	w4m	350	27	30	86
16	Los Angeles	w4w	256	23	23	7
17	New York	m4m	268	32	27	701
18	New York	m4w	284.5	34	27	984
19	New York	w4m	324.5	28	27	463
20	New York	w4w	404	33	23	131

ZOOMING IN/ZOOMING OUT: SCALES ARE MAGIC

```
var xScale = d3.scale.linear()  
    .domain([0, d3.max(data, function(d){return d.post_length;})])  
    .range([0+margin, w-margin]);  
  
var yScale = d3.scale.linear()  
    .domain([d3.max(data, function(d){return d.age;}),0])  
    .range([0+margin, h]);  
  
var rScale = d3.scale.linear()  
    .domain([0, d3.max(data, function(d) { return d.count;})])  
    .range([3,13]);
```



```
//Adjust the scales  
yScale = d3.scale.linear()  
    .domain([d3.max(data, function(d) {return d.age;}),22])  
    .range([0+margin, h]);  
  
xScale = d3.scale.linear()  
    .domain([220, d3.max(data, function(d) {return d.post_length;})])  
    .range([0+margin, w-margin]);  
  
rScale = d3.scale.linear()  
    .domain([0, d3.max(data, function(d) { return d.count;})])  
    .range([5,26]);
```

REMOVING AND ADDING CIRCLES

```
//Add 4 g groups for each of the 4 seekers types
var groupM4W=svg.append("g")
                                .attr("id", "groupM4W");
var groupM4M=svg.append("g")
                                .attr("id", "groupM4M");
var groupW4M=svg.append("g")
                                .attr("id", "groupW4M");
var groupW4W=svg.append("g")
                                .attr("id", "groupW4W");

//Creating selections of circles to append into each of the 4 seeker groups
var rmCircM4W= svg.selectAll("#m4w").remove();
var rmCircM4M= svg.selectAll("#m4m").remove();
var rmCircW4M= svg.selectAll("#w4m").remove();
var rmCircW4W= svg.selectAll("#w4w").remove();

//Removing the 4 groups from DOM and adding it back within the g group.

rmCircM4W.each(function(){
    var that = this;
    groupM4W.append(function(){ return that;});
});

rmCircM4M.each(function(){
    var that = this;
    groupM4M.append(function(){ return that;});
});

rmCircW4M.each(function(){
    var that = this;
    groupW4M.append(function(){ return that;});
});

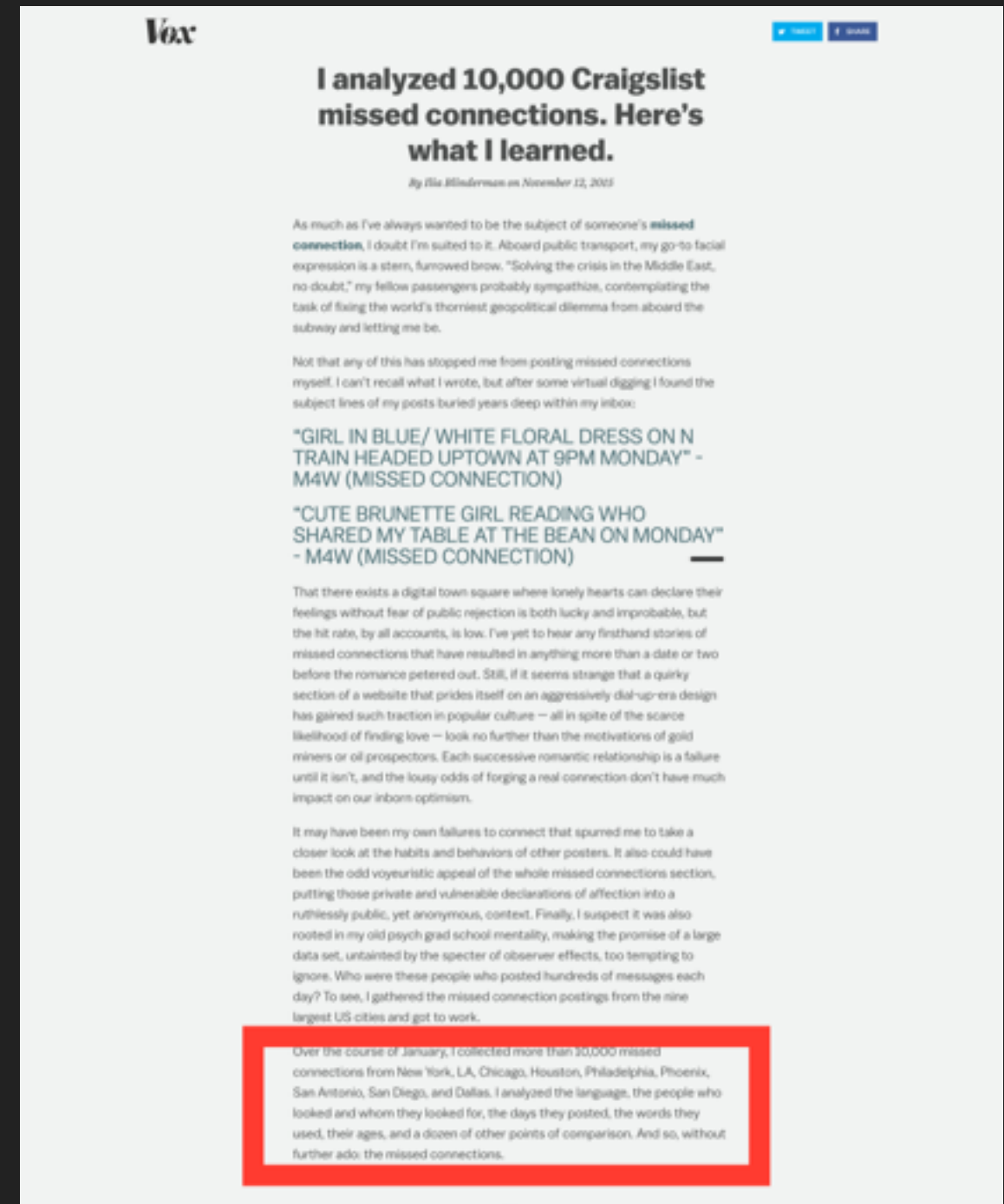
rmCircW4W.each(function(){
    var that = this;
    groupW4W.append(function(){ return that;});
});

//Adding capacity to toggle categories of seekers on and off
d3.select("rect#m4m").on("click", function(){
    var active = groupM4M.active ? false:true,
        newOpacity = active ? 0:1;

    d3.select("#groupM4M").transition().duration(200).style("opacity", newOpacity);
    groupM4M.active = active;
});
```

DISSECTING THE NARRATIVE: INTRO

- ▶ Fact: I ran a study
- ▶ Narrative: Missed Connections as a cultural phenomenon



DISSECTING THE NARRATIVE: RATES OF POSTING



- ▶ Fact: Missed connections rates, by city
- ▶ Narrative: Discussion of solitude and loneliness common to large cities



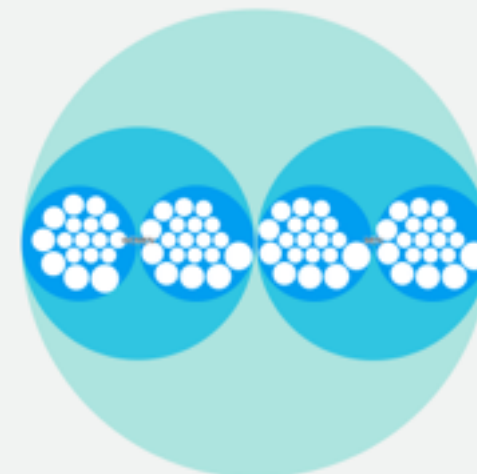
DISSECTING THE NARRATIVE: LANGUAGE USED

- ▶ Fact: Most common words
- ▶ Narrative: Discussion of the language people tend to use in online personal ads.

In an exceptionally insightful **New Yorker** piece, Nick Paumgarten, discussing online dating profiles, remarks that “demonstrating the ability, and the inclination, to write well is a rough equivalent to showing up in a black Mercedes” — partly, he says, because “males know that the best way to get laid is to send messages to as many females as possible. To be efficient, they put very little work into each message and therefore pay scant attention to each woman’s profile.” By this logic, I expected people wanting to impress a potential crush to put some hard yards into their messages; no Great American Novels, to be sure, but something witty, teasing, playful; in other words, the verbal equivalent of a black Mercedes. If

connections section, you are, by now, smirking at my naiveté. Much like in real life, the board is populated by a mixture of occasional gems filled with earnest feeling and self-reflection, a mass of posts whose allure ranges from  to  on the color scale, and a small but impassioned band of people who seek to reconnect with former neighbors from years back in hopes of finding a foot mistress or offering themselves into indentured sexual labor. Star-crossed lovers, these are not. Below, you can find the most commonly used phrases for each group (keep in mind that in a category like “women seeking women,” which had few posts to begin with, even an infrequently occurring set of words may get flagged if used, by chance, on a couple of occasions). The larger, light blue circles represent the men and women who post the connections; each contains two darker circles that represent whom they’re looking for. The white innermost circles indicate the most commonly used phrases.

Most commonly used phrases in missed connections posts



RECAP

▶ Narrative = 🔥 🔥 🔥

▶ Viz = 🔥 🔥 🔥

▶ Narrative + viz = 🎊 🔥 🎉 🎉 🎊 🎊 🎊

ADDITIONAL RESOURCES

- ▶ [Dashing D3 Newsletter](#)
- ▶ Scott Murray's [Interactive Data Visualization for the Web](#)
- ▶ Elijah Meeks' [D3.js In Action](#)