

CMPU 101 §02 · Computer Science I

Exploring Tabular Data

19 September 2022



Lab 3

Due Friday

Assignment 3

Due Wed, 11:59pm

Exam 1

Next week in lab

Review in class next Wednesday

Reminder

We won't cover everything in class!

You need to follow along with the assigned readings.

Active reading:

- Keep Pyret open and try examples.

- Take notes.

In lab and on assignments, you'll be expected to try things that may only be in the readings – or may be new altogether.

Lab and homework are additional opportunities for learning!

Where are we?

docs.google.com/spreadsheets/d/1mXH-jQw0KGQbLJnJmdfWjb

New York municipalities

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

100% \$ % .0 .00 123 Default (Ari... 10 B I S A

	A	B	C	D
1	Municipality	Kind	2010	2020
2	Adams	Town	5,143	4,973
3	Adams	Village	1,775	1,633
4	Addison	Town	2,595	2,397
5	Addison	Village	1,763	1,561
6	Afton	Town	2,851	2,769
7	Afton	Village	822	794
8	Airmont	Village	8,628	10,166
9	Akron	Village	2,868	2,888
10	Alabama	Town	1,869	1,602
11	Albany	City	97,856	99,224
12	Albion	Town	8,468	7,639
13	Albion	Town	2,073	2,009
14	Albion	Village	6,056	5,637
15	Alden	Town	10,865	9,706
16	Alden	Village	2,605	2,604
17	Alexander	Town	2,534	2,491
18	Alexander	Village	509	518
19	Alexandria	Town	4,061	3,741
20	Alexandria Bay	Village	1,078	924
21	Alfred	Town	5,237	5,157
22	Alfred	Village	4,174	4,026
23	Allegany	Town	8,004	7,493
24	Allegany	Village	1,816	1,544
25				

municipalities

Sum: 4030 Explore

```
include gdrive-sheets

include shared-gdrive("dcic-2021", "1wyQZj_L0qqV9Ekgr9au6RX2iqt2Ga8Ep")


# The ID of the Google Sheets file, which appears
# in the URL
ssid = "1mXH-jQw0KGQbLJnJmdfWjb-Ra1xf8EIwYtm3o3ajyao"

spreadsheet = load-spreadsheet(ssid)

municipalities =
  load-table:
    name :: String, kind :: String,
    pop-2010 :: Number, pop-2020 :: Number
    # true because the sheet has a "header" row
    source: spreadsheet.sheet-by-name("municipalities", true)
end
```

Now we can work with this table the same as if we'd entered it manually:

> > > **municipalities**

name	kind	pop-2010	pop-2020 
"Adams"	"Town"	5143	4973
"Adams"	"Village"	1775	1633
"Addison"	"Town"	2595	2397
"Addison"	"Village"	1763	1561
"Afton"	"Town"	2851	2769
"Afton"	"Village"	822	794
"Airmont"	"Village"	8628	10166
"Akron"	"Village"	2868	2888

What can we do with a table of data?

Get a row

Get a column in a row

Order by the values in a column

Filter by asking a question about each row

Add a column, computing a value for each row

Visualization

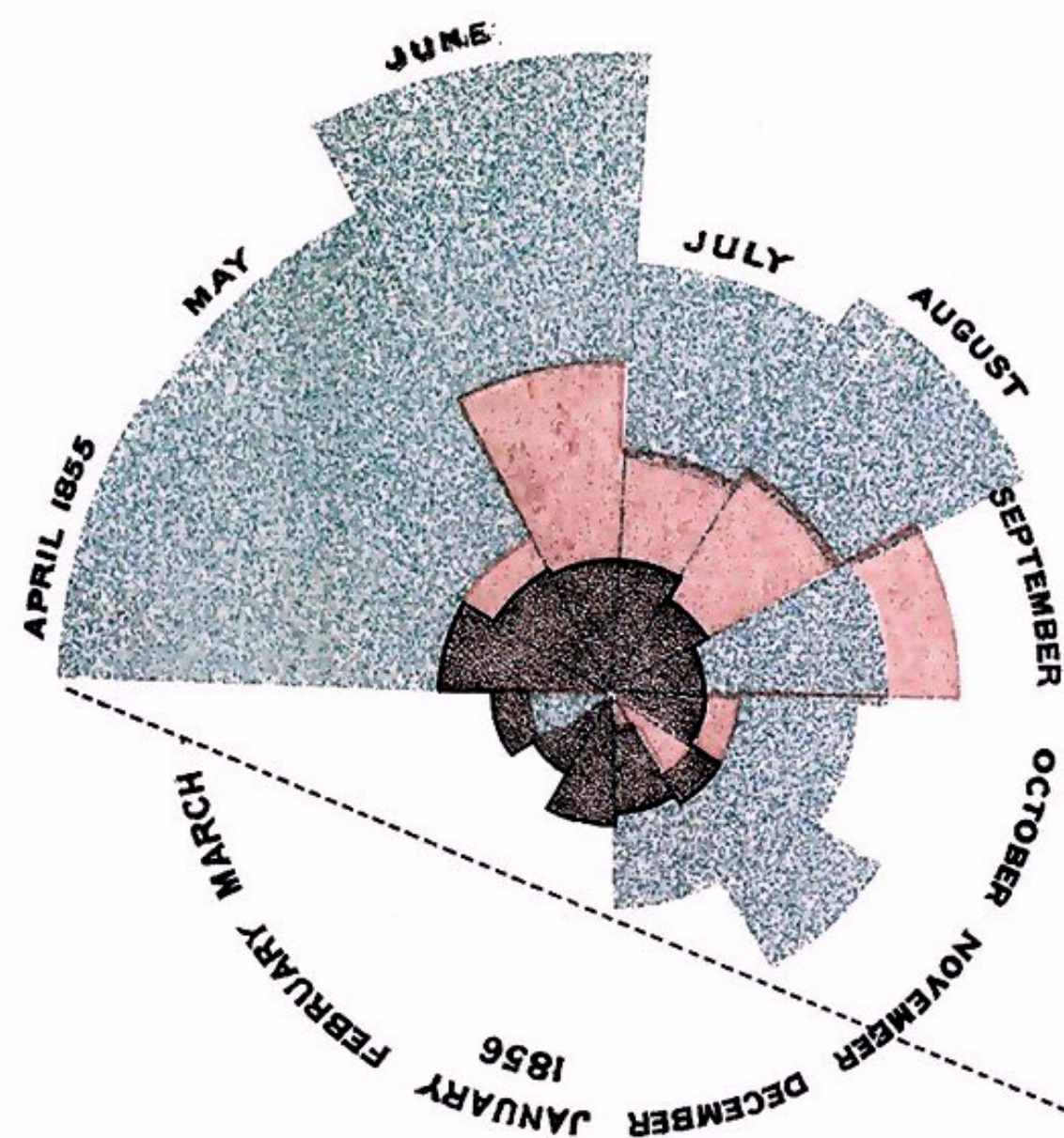
Data scientists use plots for both *exploratory* and *explanatory* purposes – they are useful for understanding data in preparation for further analysis and in presenting data to a general audience.

A visual appeal

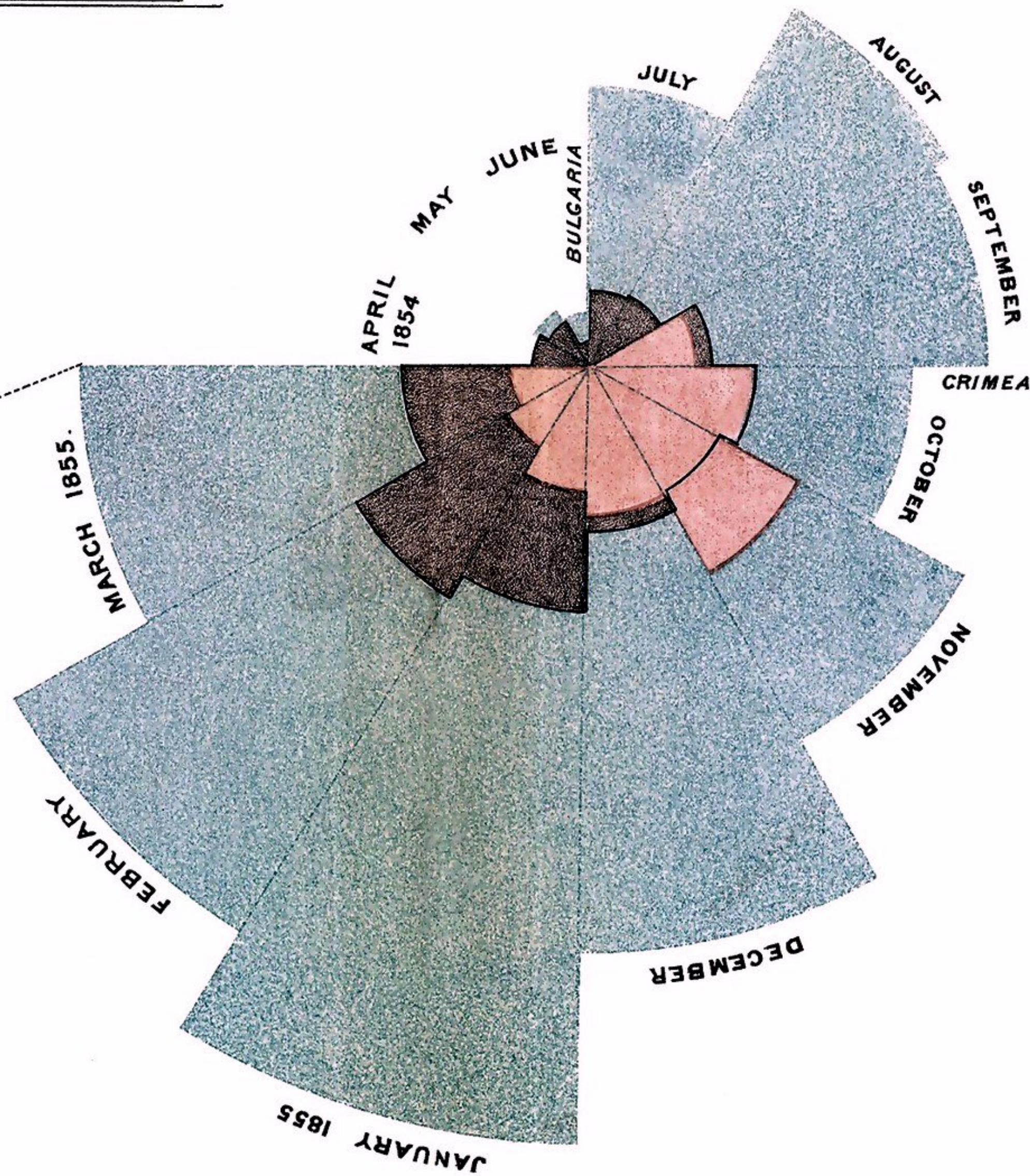
Florence Nightingale created a visualization of mortality data from the Crimean War, which was published in *Notes on Matters Affecting the Health Efficiency, and Hospital Administration of the British Army* and was sent to Queen Victoria in 1858.

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Digital visual explanations

www.pewresearch.org/next-america/#Americas-Racial-Tape

NEXT AMERICA Intro **The New Us** The Generational Divide The Showdown **PewResearchCenter**

At the same time our population is going gray, we're also becoming multi-colored. In 1960, the population of the United States was 85% white; by 2060, it will be only 43% white. We were once a black and white country. Now, we're a rainbow.

Our intricate new racial tapestry is being woven by the more than 40 million immigrants who have arrived since 1965, about half of them Hispanics and nearly three-in-ten Asians.

Because these transformations happen tick by tock, without anyone announcing them with a drum roll or press conference, they are sometimes hard to perceive.

But every so often societies experience "aha" moments, when the change is right there in plain sight. We had several such moments in early 2014, as three iconic American brands, Coke, Chevy and Cheerios, rolled out ads during the Super Bowl and Olympics that were aimed at what one voice-over called "the new us."

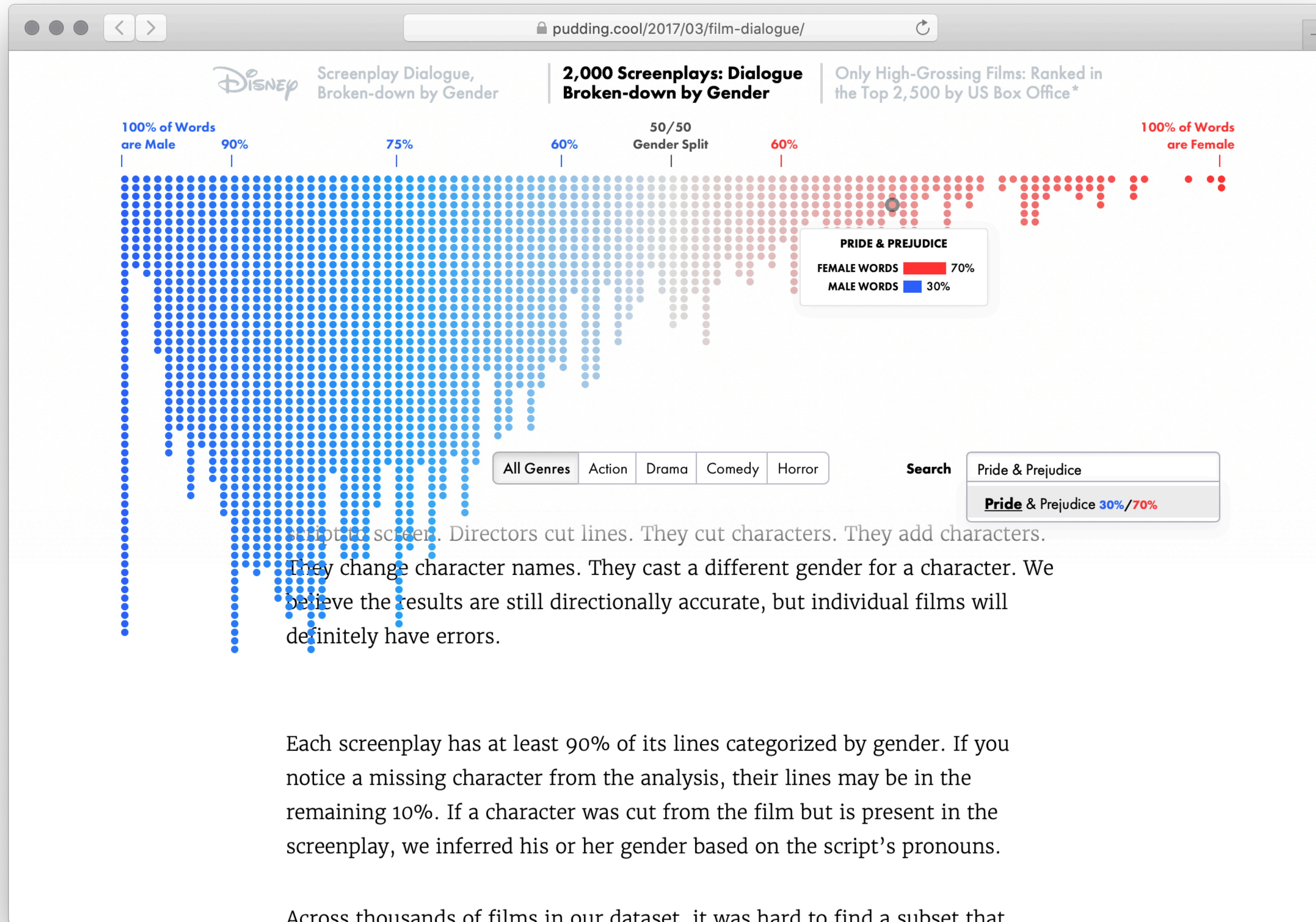
Product advertisers aren't in the business of making political statements. They're certainly not in the business of making enemies. They must have known some of their images – interracial families, same-sex parents, "America the Beautiful" sung in several

Changing Face of America
Percent of total U.S. population by race and ethnicity, 1960-2060

Year	White	Black	Hispanic	Asian	All other
1960	85%	10%	0%	0%	0%
2010	64%	12%	0%	0%	0%
2020	64%	12%	0%	0%	0%
2060	43%	13%	16%	8%	0%

Legend: All other (orange), Asian (light brown), Hispanic (dark brown), Black (gray), White (light gray)

Navigation: << >>



Home Share
spect **ARETHA FRANKLIN** • **9 to 5** **DOLLY PARTON** • Stand by Me **BEN E. KING** • Work That **MARY J. BLIGE** • High Hopes **PANIC! AT THE DISCO** • Dance to
• **SLY & THE FAMILY STONE** • Love Train **THE O'JAYS** • I'm Coming Out **DIANA ROSS** • **One Nation Under a Groove** **FUNKADELIC** • Rocking in the Free World
UNG • My Shot **LIN-MANUEL MIRANDA** • Woman **ALICIA KEYS** • Good as Hell **LIZZO** • Higher Ground **STEVIE WONDER** • Lovely Day **BILL WITHERS**
ornia Love **TUPAC SHAKUR** • **Run the Jewels** • **BEYONCÉ** • Think **ARETHA FRANKLIN** • Dis Generation **A TRIBE CALLED QUEST** • Girl on Fire **ALICIA KEYS**
olution **TOOTS AND THE MAYTALS** • Power to the People **JOHN LENNON** • Made in the USA **DEMI LOVATO** • Believer **IMAGINE DRAGONS** • Latinoamérica **CARDI B**
• God Bless the USA **LEE GREENWOOD** • **I Love Rock 'n Roll** **JOAN JETT & THE BLACKHEARTS** • Confident **DEMI LOVATO** • I Wanna Dance with Somebody (Who Loves Me)

OUR CRITICS WEIGH IN!

What Do Rally Playlists Say About the Candidates?

Presidential campaigns have a sound. We analyzed the playlists of 10 contenders to see how the songs aligned with the messages.

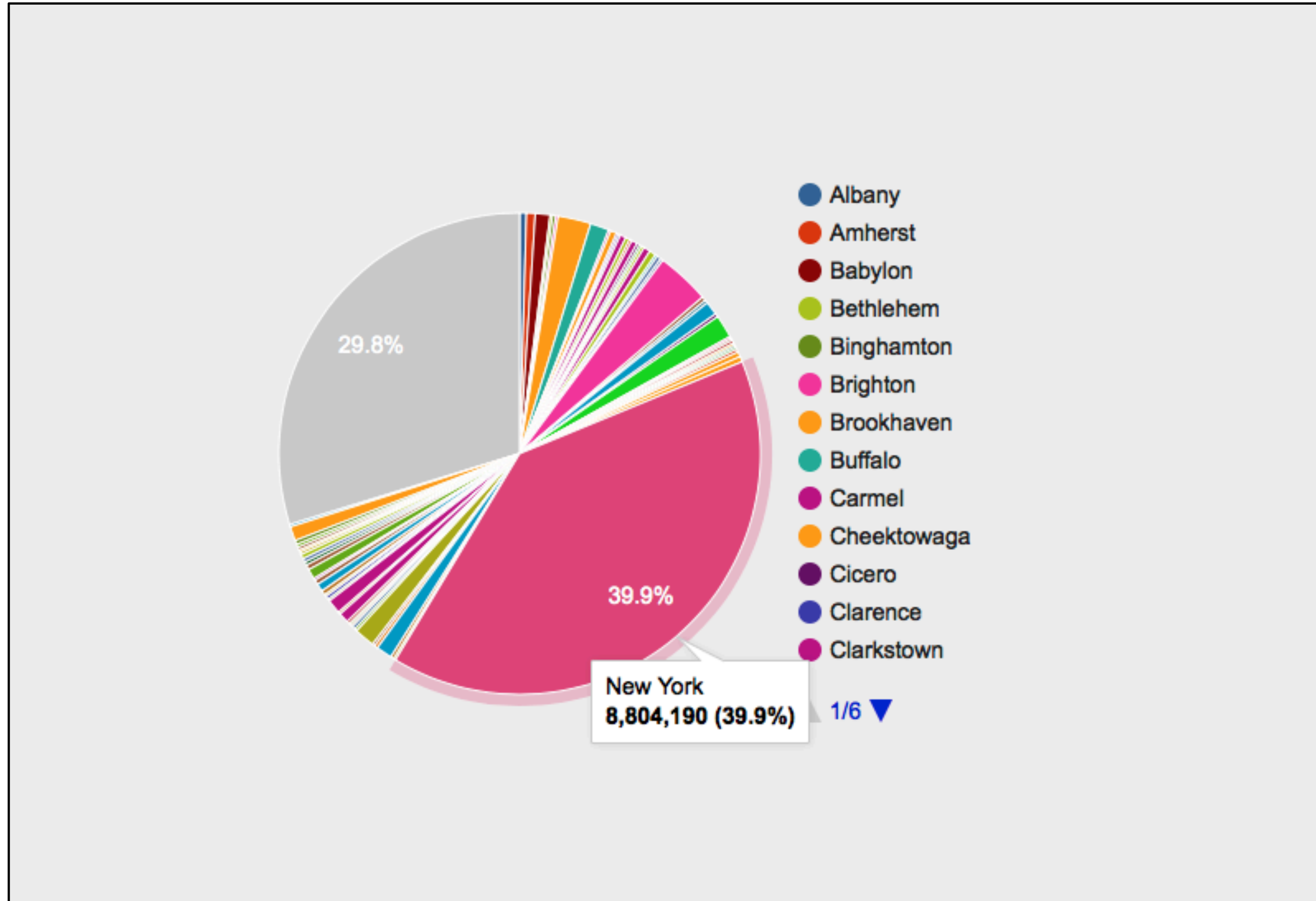
(Who Loves Me) **WHITNEY HOUSTON** • Empire State of Mind **JAY-Z (FEATURING ALICIA KEYS)** • **Dog Days Are Over** **FLORENCE + THE MACHINE** • Born This Way **LADY GAGA**
Esta Cumbia **SELENA** • Macho Man **VILLAGE PEOPLE** • Everyday People **JOAN JETT & THE BLACKHEARTS** • Clampdown **THE CLASH** • Love Train **THE O'JAYS**
Edge of Glory **LADY GAGA** • Learn to Live **DARIUS RUCKER** • Under Pressure **QUEEN** • Country Nation **BRAD PAISLEY** • Can Can **THE POINTER SISTERS**
• Baba O'Riley **THE WHO** • Love on Top **BEYONCÉ** • Soar **CHRISTINA AGUILERA** • Feeling Good **JENNIFER HUDSON** • For Something **ANDRA DAY**
FEATURING COMMON) • Fire **TIMEFLIES** • **Clampdown** **THE CLASH** • Mi Tierra **GLORIA ESTEFAN** • Let Love Rule **LENNY VAVATSIANIS** • Third Eye **FLORENCE + THE MACHINE**
HINE • The Times They Are a-Changin' **TRACY CHAPMAN** • **High Hopes** **PANIC! AT THE DISCO** • Da Da Ding **GENER8ION (FEATURING GIZZLE)** • Never Going Back
FLEETWOOD MAC • Move on Up **CURTIS MAYFIELD** • Ain't No Man **THE AVETT BROTHERS** • **Come Alive** **YEARS & YEARS WITH JESS GLYNNE** • Brooklyn Go Home
• On the Road Again (Live) **WILLIE NELSON** • Revolution **TOOTS AND THE MAYTALS** • Uprising **MUSE** • Revolution **FLOGGING MOLLY** • The Revolution Starts Now
VE EARLE • Power to the People **JOHN LENNON** • Takin' It to the Streets **THE DOOBIE BROTHERS** • Heroes **DAVID BOWIE** • All We Ever Knew **THE HEAD AND THE TALES**

TURN YOUR SOUND ON!

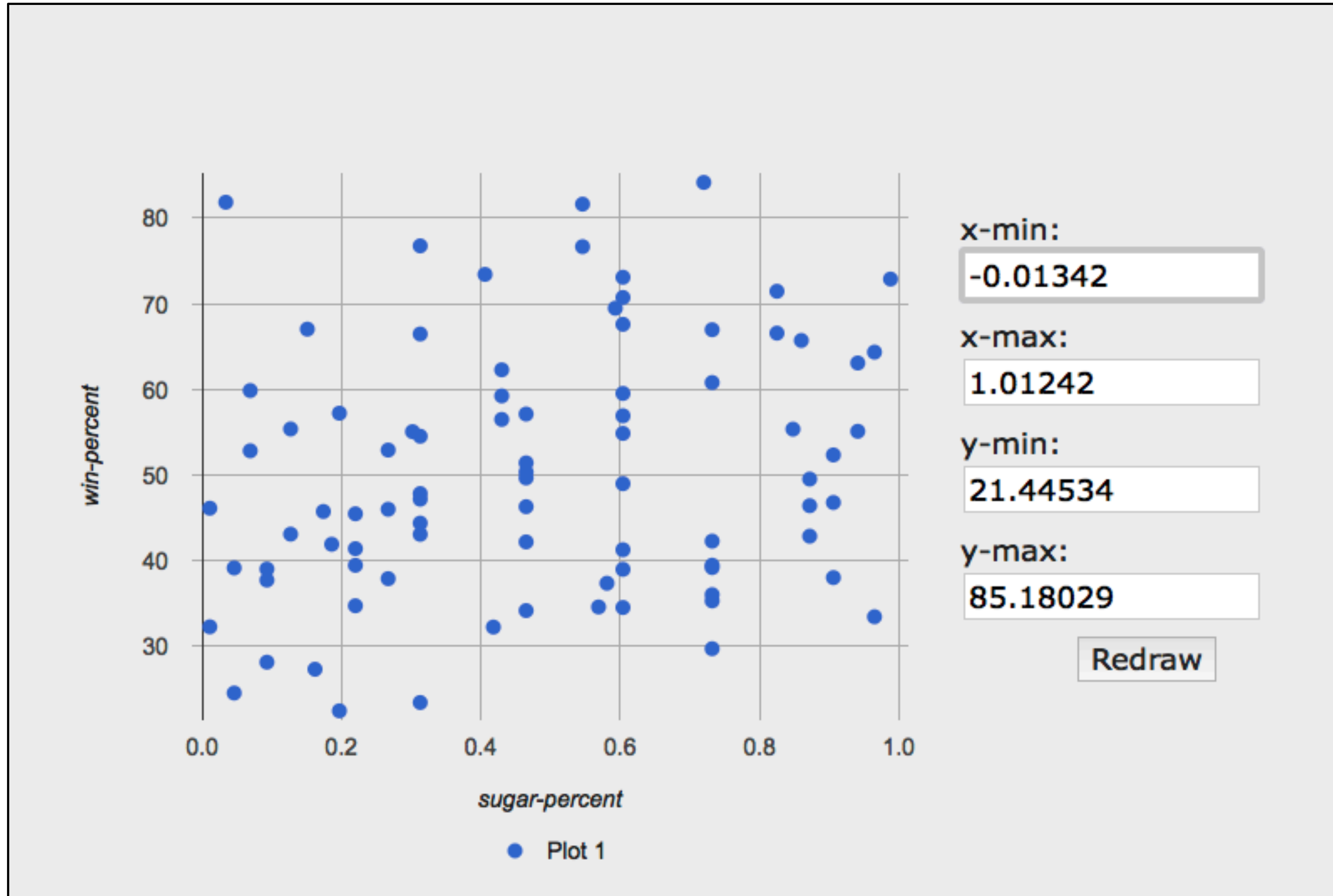
The **dcic-2021** library we've been using to work with tables includes several functions to generate different kinds of plots like the ones we've talked about.

How is population distributed in the state?
pie-chart(municipalities, "name", "pop-2020")

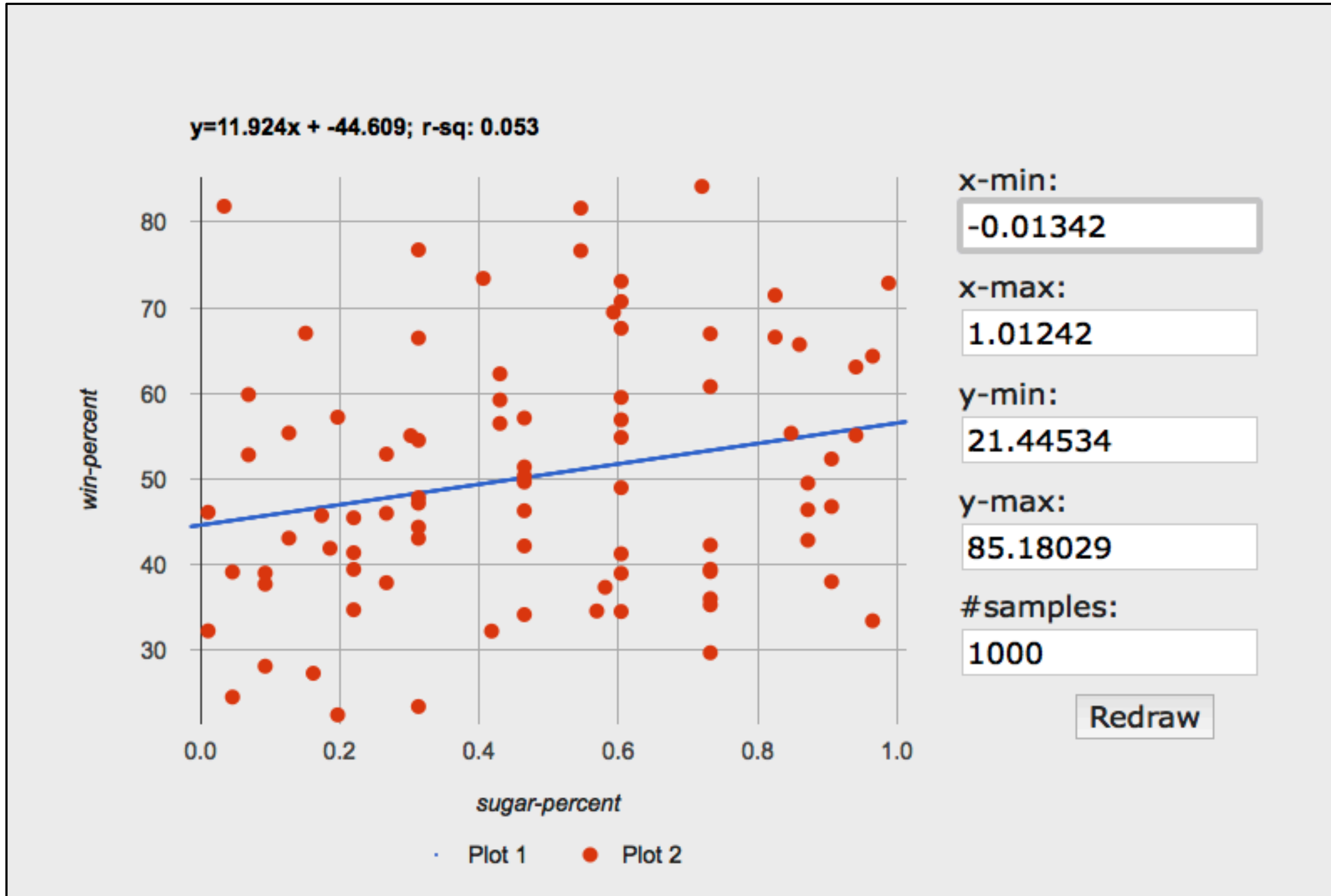
How is population distributed in the state?
pie-chart(municipalities, "name", "pop-2020")



```
scatter-plot(candy-data,  
            "sugar-percent", "win-percent")
```

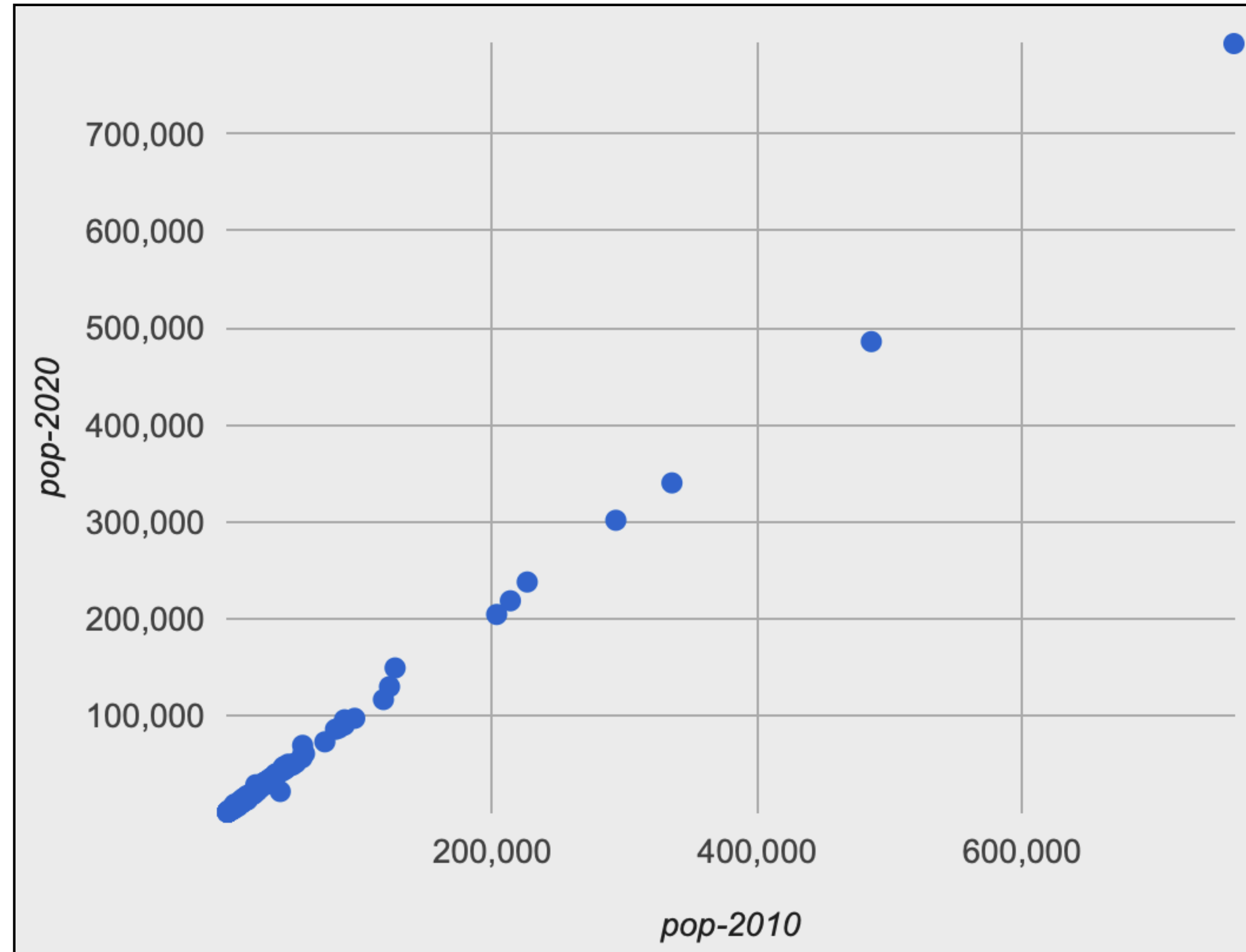



```
lr-plot(candy-data,  
        "sugar-percent", "win-percent")
```



*# Is a town's population in 2010 correlated with
its population in 2020?*

```
scatter-plot(ft, "pop-2010", "pop-2020")
```



```
ft = fastest-growing-towns(municipalities)
```

```
# Visually present the growth data
```

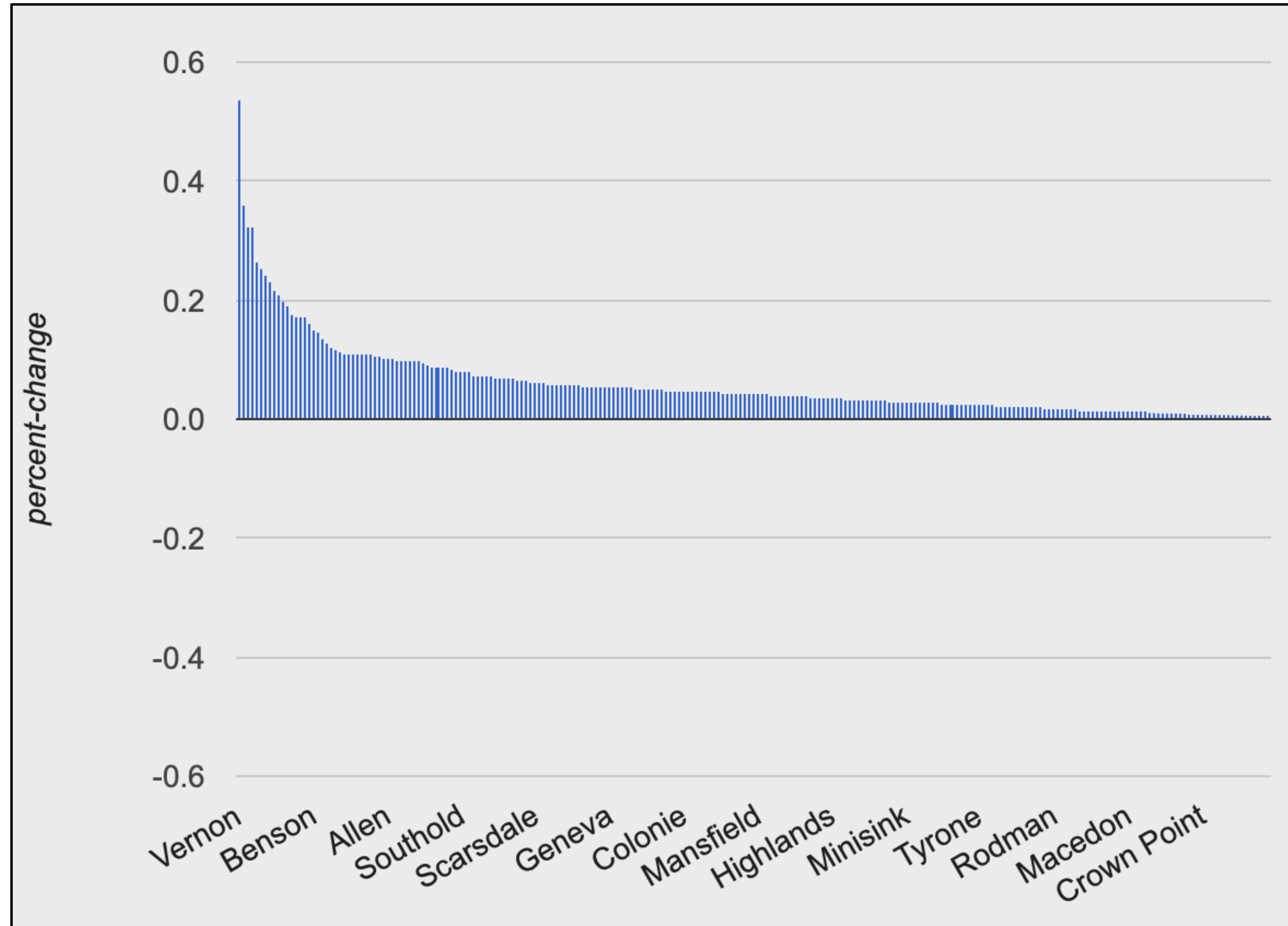
```
bar-chart(ft, "name", "percent-change")
```



```
ft = fastest-growing-towns(municipalities)
```

```
# Visually present the growth data
```

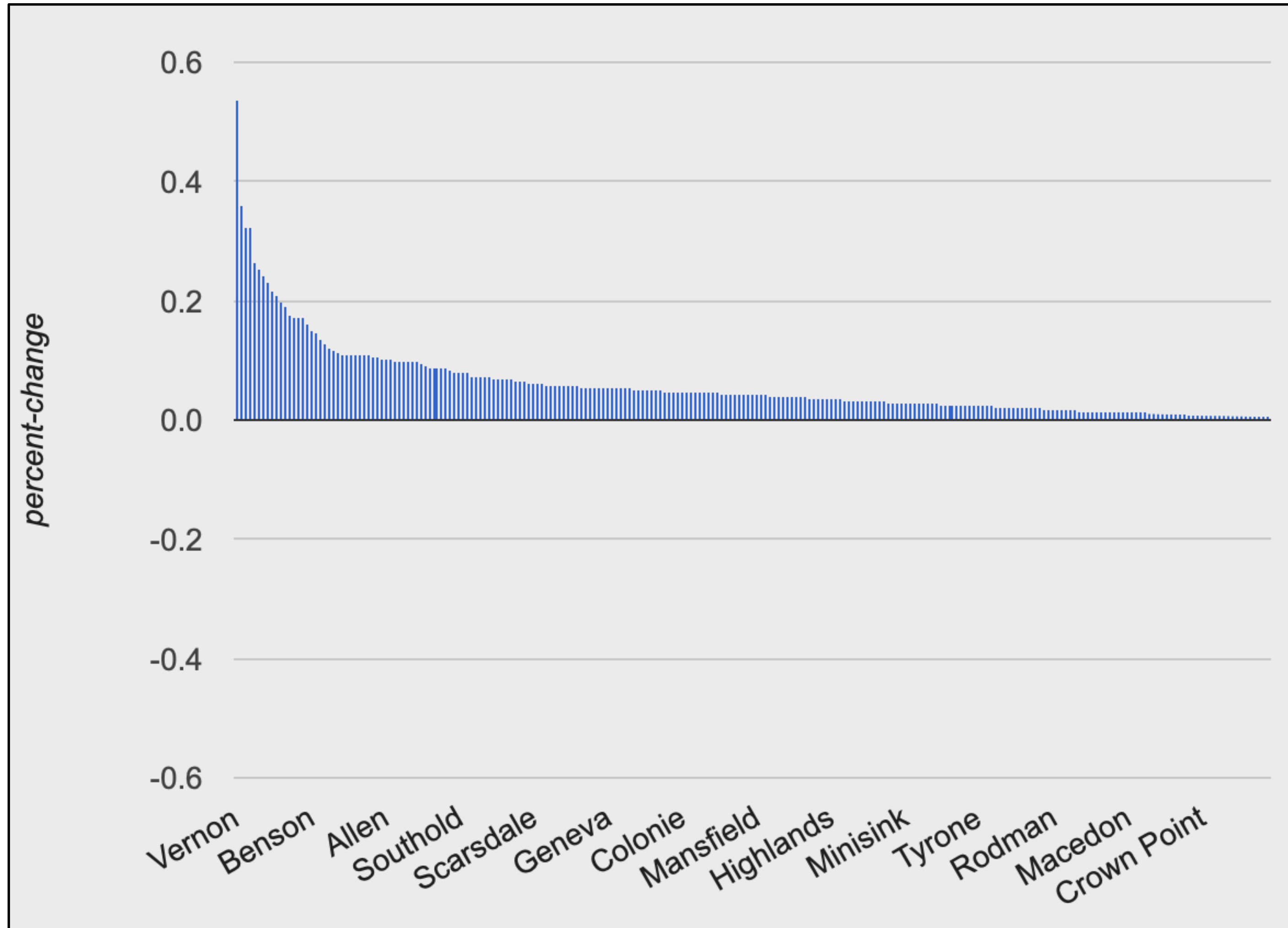
```
bar-chart(ft, "name", "percent-change")
```



```
ft = fastest-growing-towns(municipalities)
```

```
# Visually present the growth data
```

```
bar-chart(ft, "name", "percent-change")
```

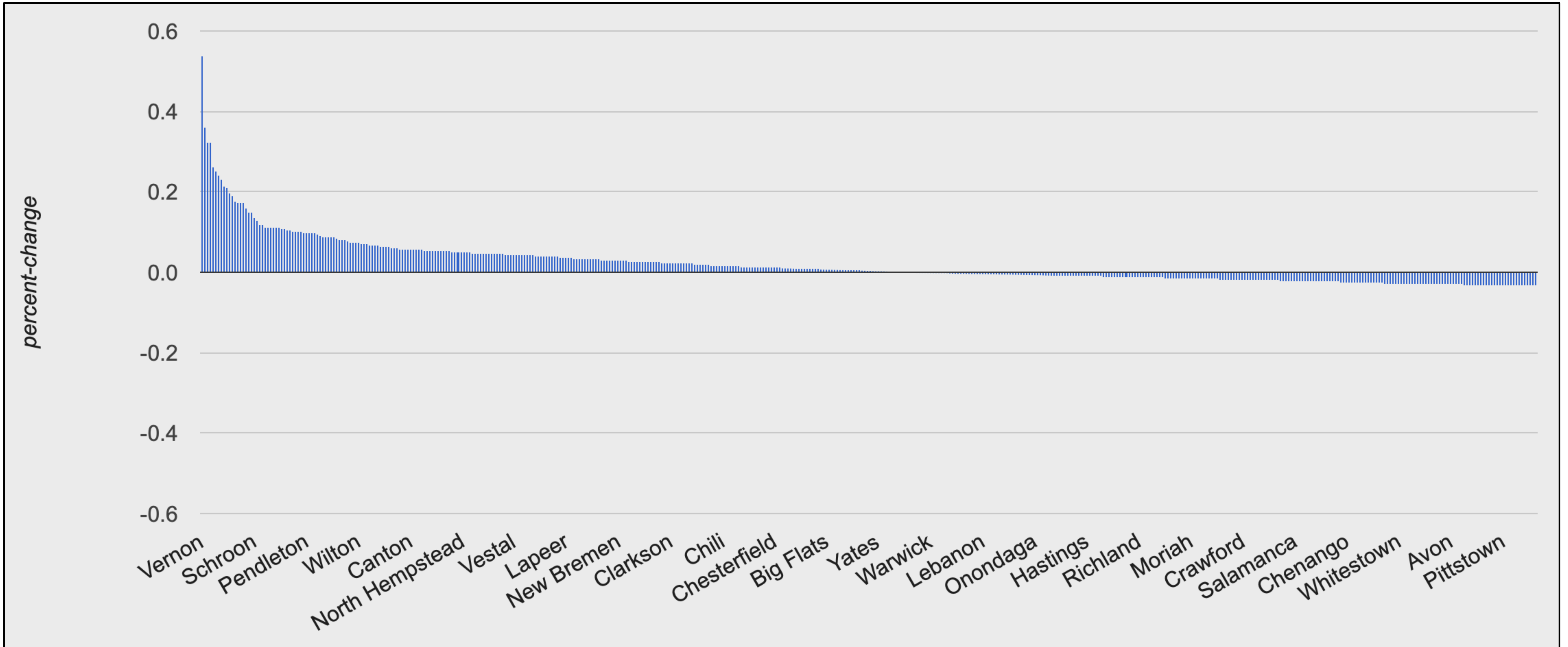


...didn't any towns shrink?

ft = fastest-growing-towns(municipalities)

Visually present the growth data

bar-chart(*ft*, "name", "percent-change")



```
ft = fastest-growing-towns(municipalities)
```

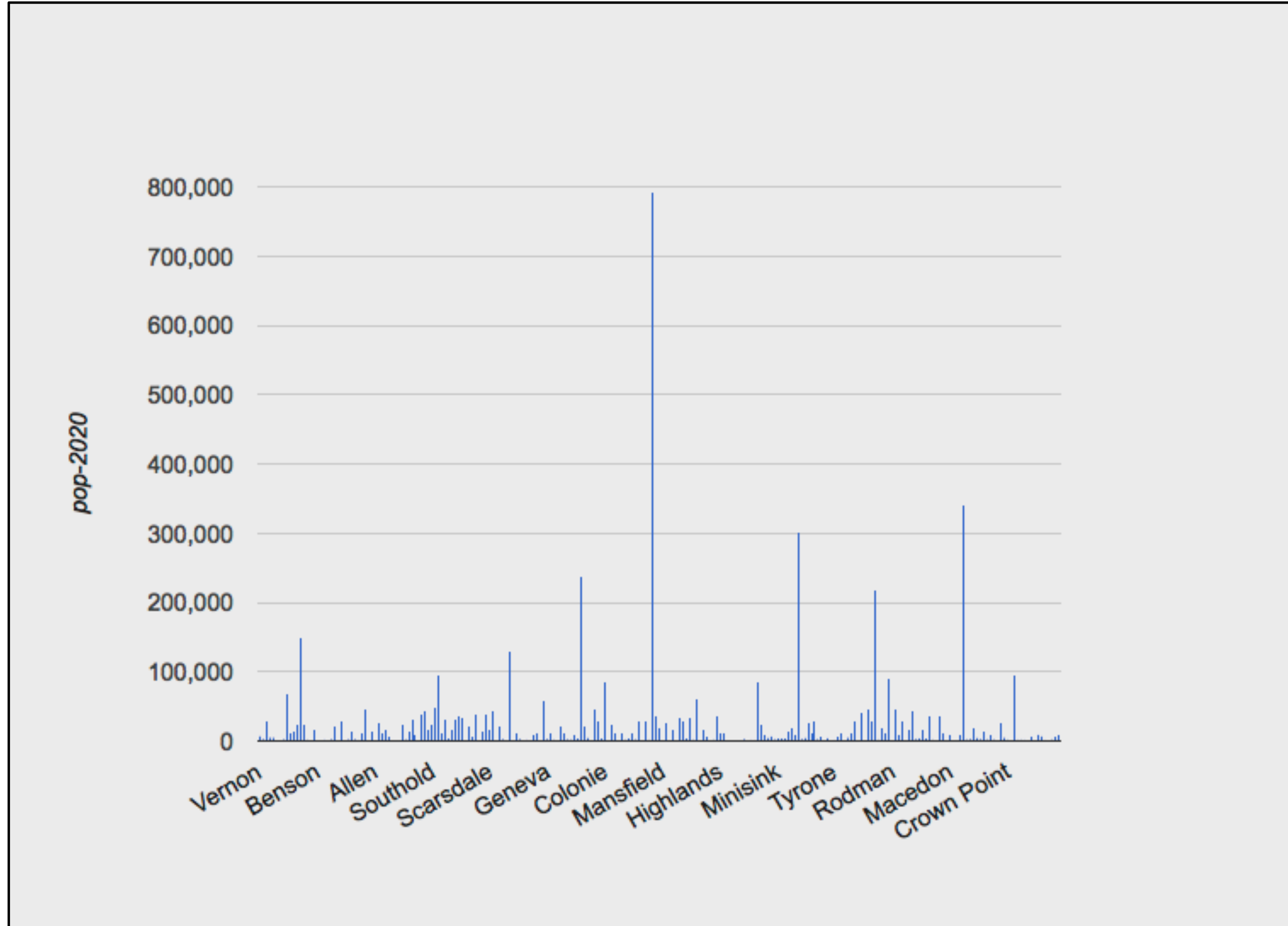
```
# Visually present the growth data
```

```
bar-chart(ft, "name", "pop-2020")
```

ft = fastest-growing-towns(municipalities)

Visually present the growth data

bar-chart(*ft*, "name", "pop-2020")



Fox in bedroom,
dog trapped in wall

There are a staggering number of publicly available data sets that we can load from a spreadsheet.

Take a look at the archives of the Data is Plural newsletter:

data-is-plural.com

But real data sets can be significantly harder to work with than contrived examples:

- Missing values

- Inconsistent entry of data

- Differing levels of precision (1987 vs 7 July 1987)

We can use Pyret to explore these data sets and transform them so they're easier for us to work with.



LONDON FIRE BRIGADE

Animal rescue incidents attended by LFB

London Fire Brigade

Data

Created 6 years ago, updated 18 days ago

The London Fire Brigade attends a range of non-fire incidents (which we call 'special services'). These 'special services' include assistance to animals that may be trapped or in distress.

We routinely get asked for information about the number of such incident attended by the London Fire Brigade and this data is published on the London Datastore to assist those who require it.

The data is provided from January 2009 and is updated monthly. A range of information is supplied for each incident including some location information (postcode, borough, ward), as well as the data/time of the incidents. We do not routinely record data about animal deaths or injuries.

Please note that any cost included is a notional cost calculated based on the length of time rounded up to the nearest hour spent by Pump, Aerial and FRU appliances at the incident and charged at the current Brigade hourly rate.

The London Fire Commissioner is the fire and rescue authority for London and runs the London Fire Brigade.


- LFB London Fire Brigade
- LFB
- Emergency response
- rescue
- special services
- animals
- animal


Animal Rescue incidents attended by LFB from Jan 2009.csv (2.79 MB)

From	To
01/01/2009	31/12/2021

Preview
Download

Author


[LFB Information Management](#)

This data is available as CSV – a plain-text file where each cell of the spreadsheet is separated by commas.

To load it into Pyret, we can first upload it to a Google spreadsheet.

docs.google.com/spreadsheets/d/1UdddiozfEJ0U0GntBpyQ4zZncBnP

Untitled spreadsheet ☆

File Edit View Insert Format Data Tools Extensions Help

100% \$ % .0 .00 123 Default (Ari... 10 B I S A

A1 fx

A B C D E F G H I J K

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Sheet1

Import file

File

Animal Rescue incidents attended by LFB from Jan 2009.csv

Import location: Replace spreadsheet

Separator type: Detect automatically

Convert text to numbers, dates, and formulas

Import data Cancel

docs.google.com/spreadsheets/d/1UdddiozfeJ0U0GntBpyQ4zZncBnP

Animal rescue

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

100% \$ % .0 .00 123 Arial 10 B I S A

	A	B	C	D	E	F	G	H	I	J	K
1	IncidentNumber	DateTimeOfCall	CalYear	FinYear	TypeOfIncident	PumpCount	PumpHoursTotal	HourlyNotionalCo	IncidentNotionalC	FinalDescription	AnimalGroupPa
2	139091	01/01/2009 3:01	2009	2008/09	Special Service	1	2	255	510	Redacted	Dog
3	275091	01/01/2009 8:51							255	Redacted	Fox
4	2075091	04/01/2009 10:07							255	Redacted	Dog
5	2872091	05/01/2009 12:27							255	Redacted	Horse
6	3553091	06/01/2009 15:20							255	Redacted	Rabbit
7	3742091	06/01/2009 19:30							255	Redacted	Unknown - Heav
8	4011091	07/01/2009 6:29	2009	2008/09	Special Service	1	1	255	255	Redacted	Dog
9	4211091	07/01/2009 11:55							255	Redacted	Dog
10	4306091	07/01/2009 13:48							255	Redacted	Squirrel
11	4715091	07/01/2009 21:24							255	Redacted	Dog
12	5186091	08/01/2009 14:34							255	Redacted	Cat
13	5363091	08/01/2009 19:20							510	Redacted	Bird
14	5682091	09/01/2009 11:01							255	Redacted	Dog
15	5688091	09/01/2009 11:18							255	Redacted	Dog
16	5724091	09/01/2009 12:40							255	Redacted	Cat
17	5770091	09/01/2009 13:43							255	Redacted	Cat
18	5789091	09/01/2009 14:30							255	Redacted	Cat
19	5797091	09/01/2009 14:39							255	Redacted	Dog
20	6259091	10/01/2009 9:35							255	Redacted	Unknown - Dom
21	6270091	10/01/2009 10:09	2009				1	255	255	Redacted	Dog
22	6317091	10/01/2009 11:27	2009				1	255	255	Redacted	Cat
23	6353091	10/01/2009 12:26	2009				1	255	255	Redacted	Cat
24	6355091	10/01/2009 12:32	2009				1	255	255	Redacted	Dog
25	6378091	10/01/2009 13:09	2009				1	255	255	Redacted	Cat
26	6400091	10/01/2009 13:28	2009	2008/09	Special Service	1	1	255	255	Redacted	Dog
27	6530091	10/01/2009 16:38	2009	2008/09	Special Service	1	1	255	255	Redacted	Cat

Share with people and groups

No one has been added yet

Get link

<https://docs.google.com/spreadsheets/d/1UdddiozfeJ0U0GntBpyQ4zZnc...> [Copy link](#)

Restricted

- Restricted
- Vassar Google Apps for Education
- Anyone with the link

[Send feed](#) [Done](#)

Animal Rescue incidents attended by LFB from Jan 2009

Explore

```
include gdrive-sheets
```

```
include shared-gdrive("dcic-2021", "1wyQZj_L0qqV9Ekgr9au6RX2iqt2Ga8Ep")
```

```
# The ID of the Google Sheets file, which appears in the URL
```

```
ssid = "1i8qHPZC2mHRcfPRwBnSZRLpe0GEnZXmAo0cAh-fc3mk"
```

```
spreadsheet = load-spreadsheet(ssid)
```

```
rescue-data =
```

```
  load-table:
```

```
    ...
```

```
    source: spreadsheet.sheet-by-name("rescues", true)
```

```
end
```

Too many columns!

We could copy-and-paste the names to have our column names in Pyret, but, instead, let's trim columns we don't care about first.


```
include gdrive-sheets
```

```
include shared-gdrive("dcic-2021", "1wyQZj_L0qqV9Ekgr9au6RX2iqt2Ga8Ep")
```

```
# The ID of the Google Sheets file, which appears in the URL
```

```
ssid = "1i8qHPZC2mHRcfPRwBnSZRLpe0GEnZXmAo0cAh-fc3mk"
```

```
spreadsheet = load-spreadsheet(ssid)
```

```
rescue-data =
```

```
  load-table:
```

```
    time, year, description, animal-group-parent, origin-of-call,
```

```
    property-type, property-category, special-service-type-category,
```

```
    special-service-type, ward, borough, stn-name, street, postcode
```

```
    source: spreadsheet.sheet-by-name("rescues-simple", true)
```

```
end
```

	ward	bor
tion"	some("Crystal Palace & Upper Norwood")	some
tion"	some("Woodside")	some
	some("Carshalton Central")	some
	some("Harefield")	some
tion"	some("Gooshays")	some
tion"	some("Alibon")	some)
tion"	some("Cathall")	some
	some("Wanstead")	some
	some("New Addington North")	some
	some("Lea Bridge")	some

To make the data end up in the format we want, we'll use *sanitizers*, which convert data from an external source into a specific Pyret data type.

Built-in sanitizers:

string-sanitizer

Replaces missing values with ""

Converts non-string data to strings, e.g., 3 to "3"

num-sanitizer

Replaces missing values with 0

Converts numeric strings to numbers, e.g., "3" to 3

Sanitizers are just functions, so you can write your own too!

To be continued on Wednesday!

Class code:

https://code.pyret.org/editor#share=16D8g-f5VwpUqHnm_lHKej_lIG1A2Hz-NvI&v=31c9aaf