

Web Scraping

The Connecticut Lottery

Neal Drakos
May 2017

Methodology

- Scrap the data from the State of CT Lottery website
- data cleaning and analysis in R to look for:
- Winning trend over the years. Are people winning more or less often?
- Location- Where are the most winning tickets sold?
- Games- Which are the highest paying / most frequent amount of winning games?

Scrappy

The screenshot displays the CT Lottery KENO website. At the top, the CT Lottery logo is on the left, and the word "KENO" in large, colorful letters is in the center. To the right of "KENO" are several large numbers: 13, 41, and 27, along with social media icons for Facebook, Twitter, YouTube, and a mobile app. Below the header is a navigation bar with links: HOME, ABOUT US, WHERE THE MONEY GOES, GAMES & WINNING NUMBERS, WINNERS (highlighted with a red box), CLAIM A PRIZE, BE A RETAILER, and CORPORATE RESPONSIBILITY. A search bar is also present.

Below the navigation bar, there are two main sections. On the left, under "WINNING NUMBERS", there is a list of recent drawings with dates and winning numbers. On the right, under "LOTTERY", there is a dropdown menu with options: WINNERS (highlighted with a red box), WINNER STORIES, VIDEOS, and WINNER PUBLICITY.

The "WINNING NUMBERS" section lists the following drawings:

Game	Date	Winning Numbers	Check Lines	Payout
LUCKY LINKS	May 9, 2017	Day 1-6-12-13-16-18-20-22	Check Lines	Payout
	May 8, 2017	Night 1-2-5-8-11-12-13-22	Check Lines	Payout
POWERBALL	May 6, 2017	11-21-31-41-59 PB-21 PP-3	Check Lines	Payout
MEGA MILLIONS	May 5, 2017	4-23-33-47-53 MB-7 MP-4	Check Lines	Payout
10to1	May 5, 2017	3-4-5-11-30-33	Check Lines	Payout
LUCKY 75	May 8, 2017	4-7-23-39-48 LB-2	Check Lines	Payout
LUCKY 75	May 8, 2017		Check Lines	Payout

The "LOTTERY" section features two large promotional banners. The left banner is for "30 GRAND" with a prize of "WIN UP TO \$30,000!" and "WIN UP TO 15 TIMES!". The right banner is for "MONEY" with a prize of "WIN UP TO \$6,500,000 IN TOTAL GAME PRIZES!" and "WIN UP TO 20 TIMES!". Both banners show "WINNING NUMBERS" and "YOUR NUMBERS" sections.

At the bottom of the page, there is a footer with the text "Multiply the fun!" and a date range "MAY 8-15".

About the data

- Approximately 2,000 rows of winners of the CT lottery for prizes \$10,000 and above
- dates: 2008- April 2017^{***}Excluding 2010
- Variables:

Date	Winner	Retailer	Retailer location	Game	Prize
(purchase date)	(name & hometown)	(store name)	(town)	(name of game)	(amount won)

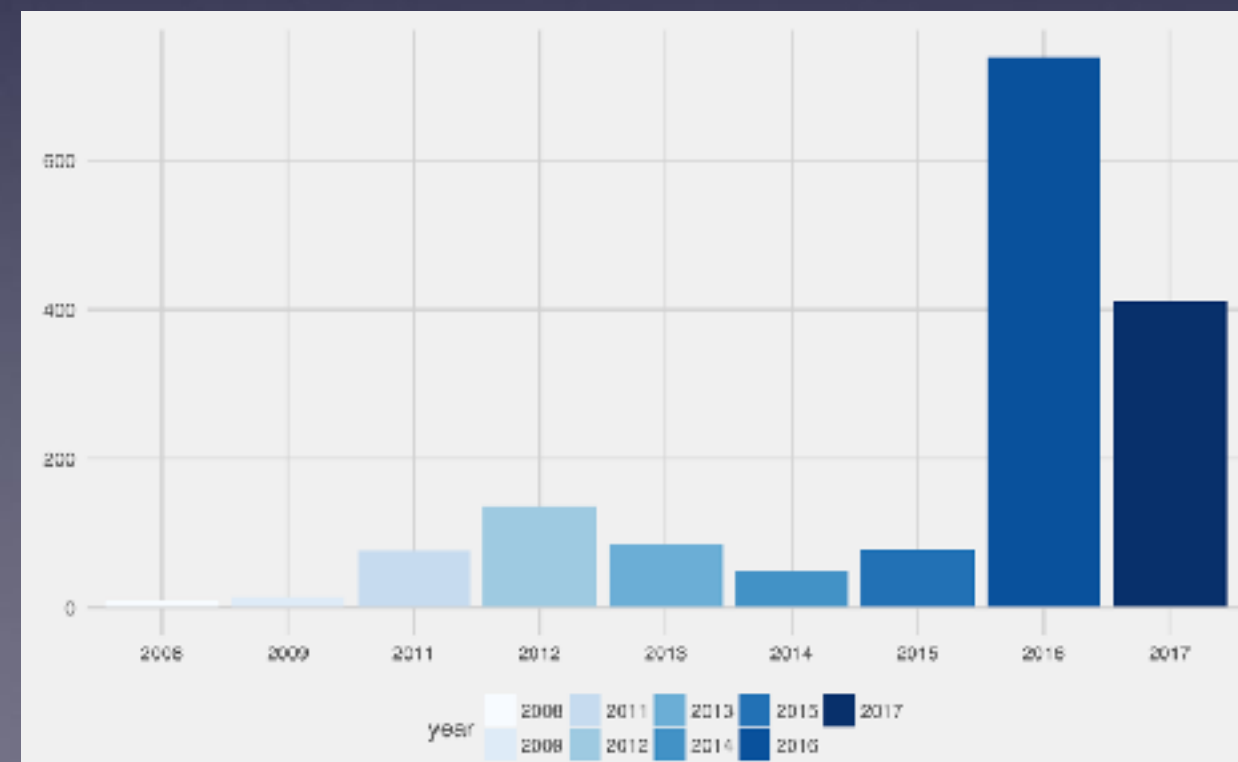
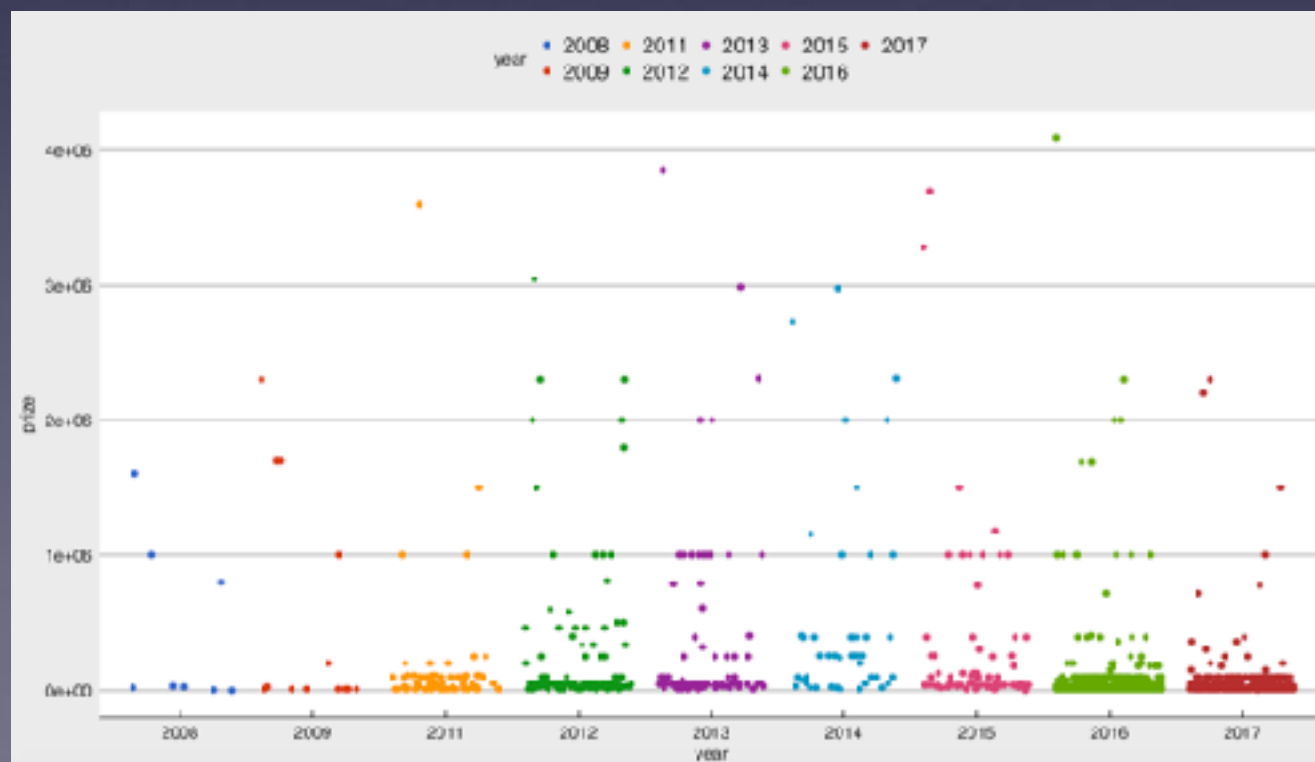
Overview

Total prize money by year

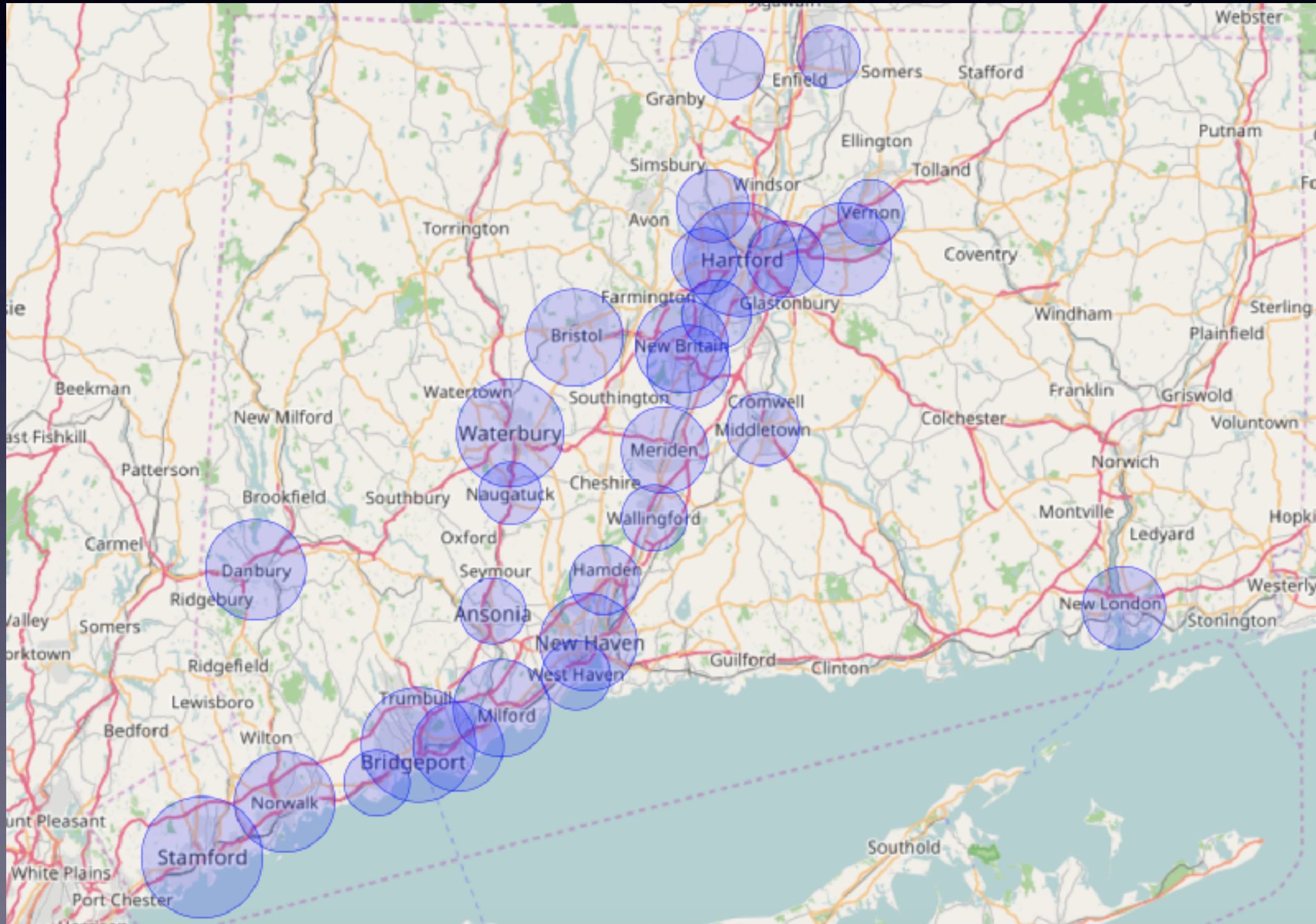
	year	total
1	2011	276193678
2	2012	91933400
3	2016	46620765
4	2009	41995000
5	2013	30442094
6	2014	25637940
7	2017	22875811
8	2015	22321672
9	2008	17780650

Winners count by year

	year	number
1	2016	738
2	2017	411
3	2012	137
4	2013	85
5	2015	78
6	2011	76
7	2014	49
8	2009	15
9	2008	10

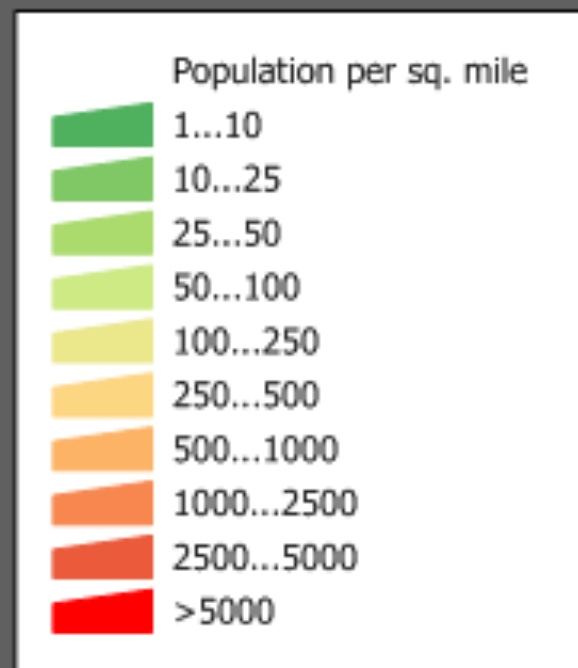


Top locations that sell the most amount of winning tickets over 10k

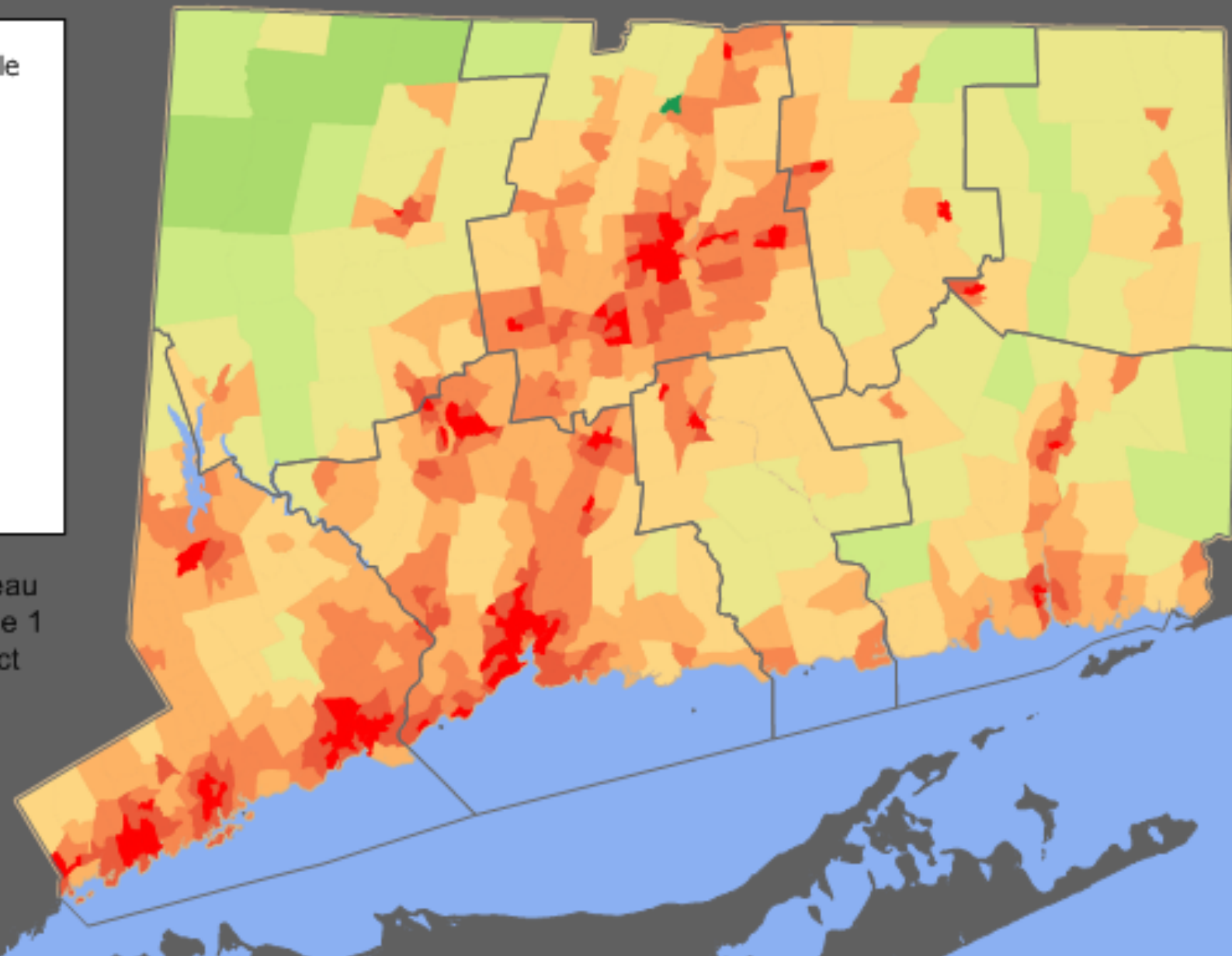


	retailer_location	count	lon	lat
1	Stamford, CT	74	-73.53873	41.05343
2	Bridgeport, CT	65	-73.19518	41.18655
3	Hartford, CT	63	-72.68509	41.76371
4	Waterbury, CT	54	-73.05150	41.55815
5	Danbury, CT	49	-73.45401	41.39482
6	Norwalk, CT	49	-73.40816	41.11774
7	Milford, CT	47	-73.05404	41.23070
8	Bristol, CT	46	-72.94927	41.67176
9	New Haven, CT	46	-72.92788	41.30827
10	Manchester, CT	41	-72.52150	41.77593
11	Stratford, CT	41	-73.13317	41.18454
12	New Britain, CT	40	-72.77954	41.66121
13	Meriden, CT	35	-72.80704	41.53815
14	Groton, CT	33	-72.07907	41.34975
15	Kensington, CT	32	-72.75871	41.63538
16	East Hartford, CT	28	-72.61283	41.76342
17	Middletown, CT	26	-72.65063	41.56232
18	Bloomfield, CT	25	-72.73009	41.82649
19	West Haven, CT	24	-72.94697	41.27055
20	Hamden, CT	23	-72.90261	41.38388

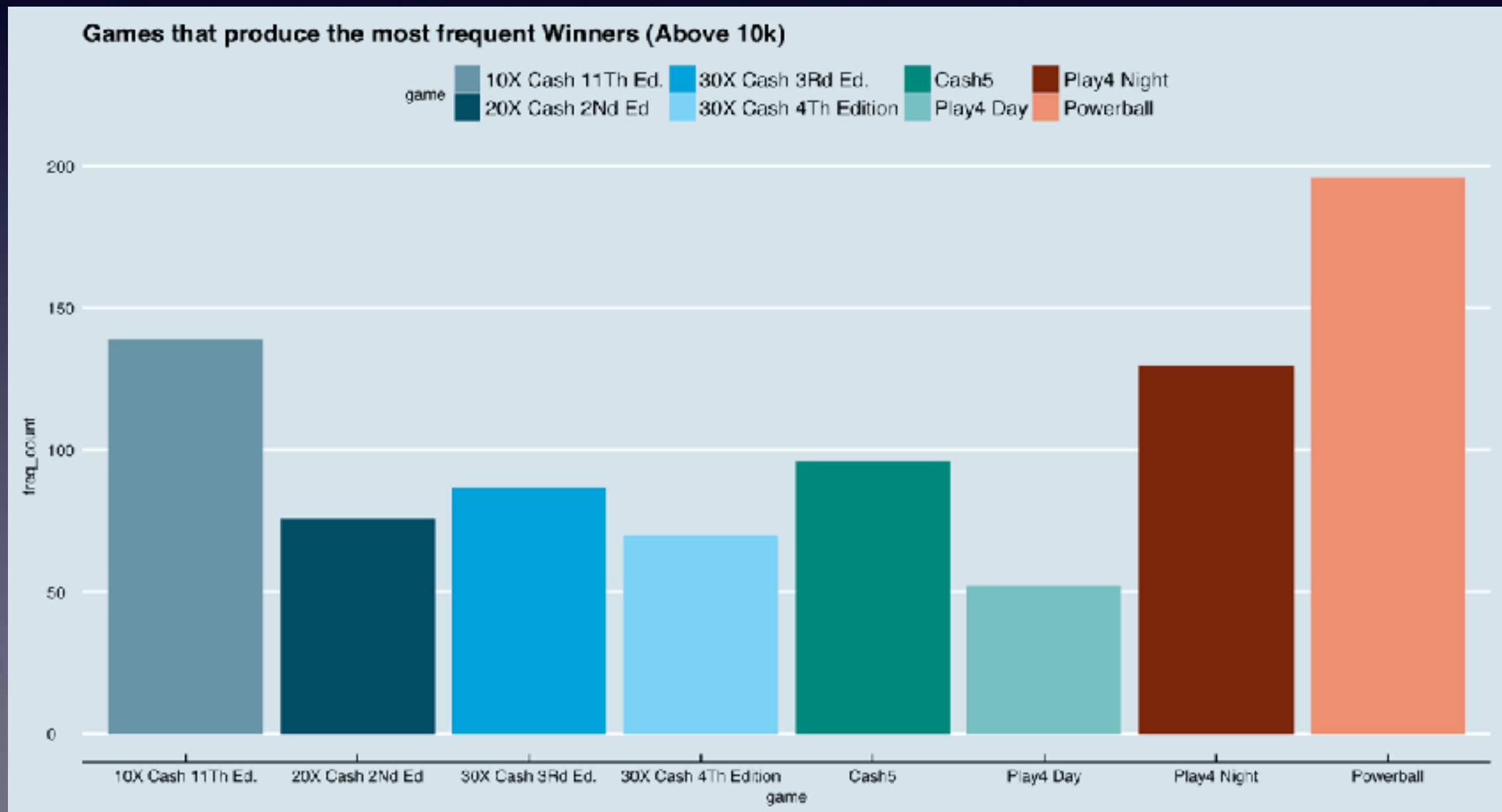
Population density



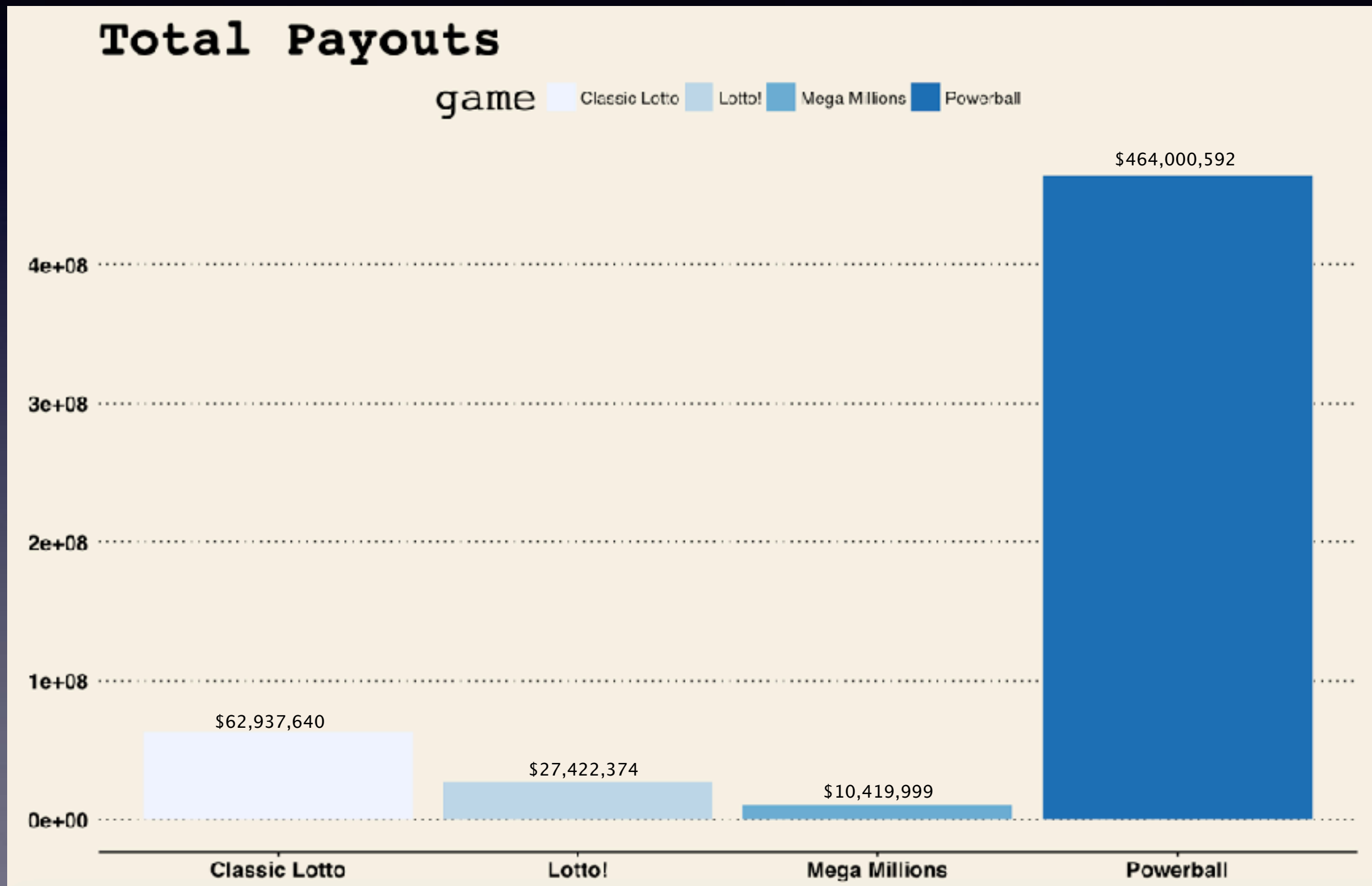
Source: U.S. Census Bureau
Census 2010 Summary File 1
population by census tract



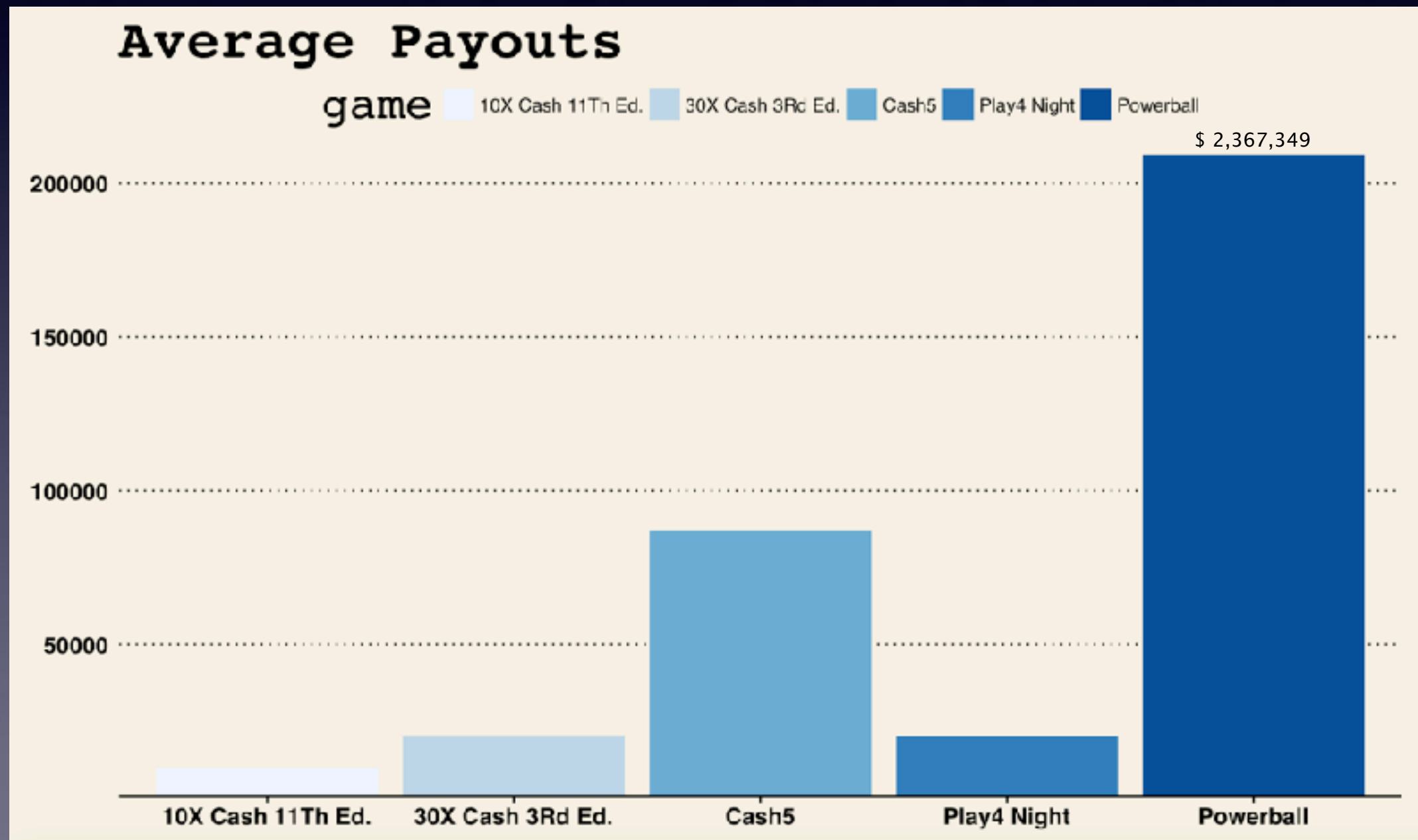
Games



Games that pay



Average Payouts for top games (2008-Present)



Winners by Month

