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?

Scrapy

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
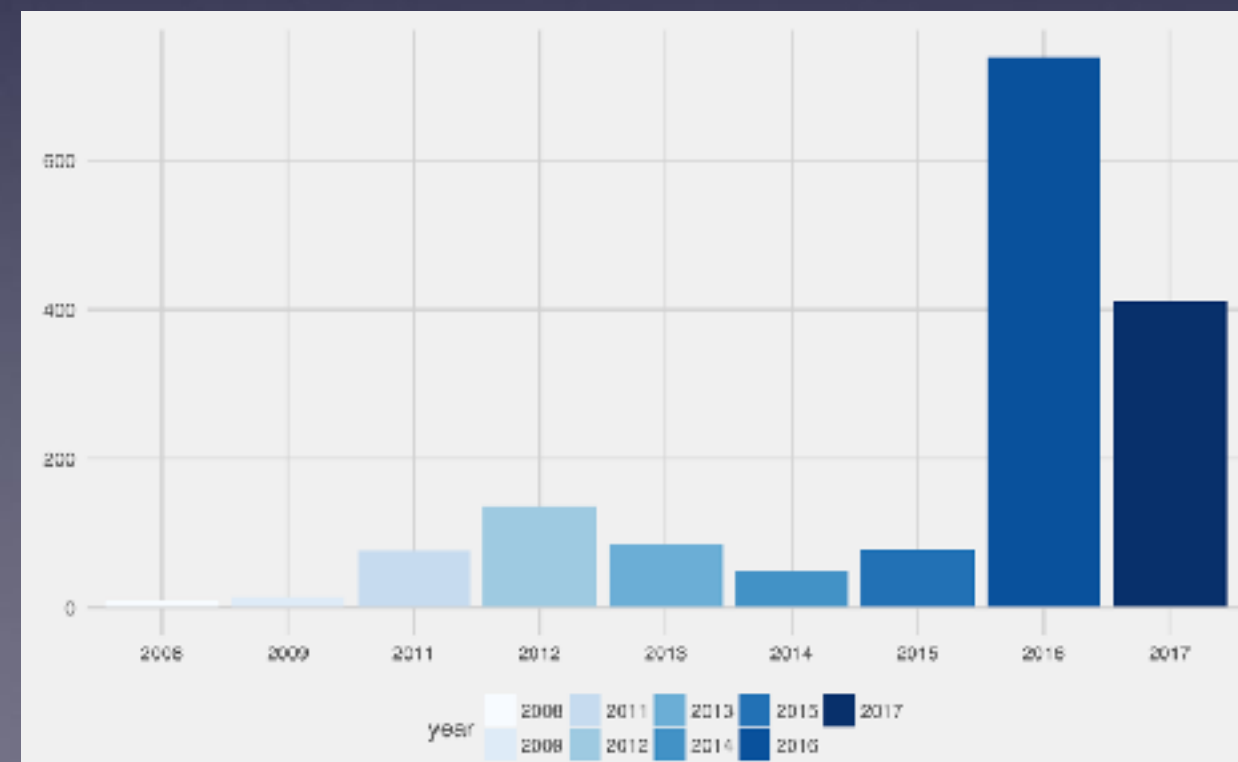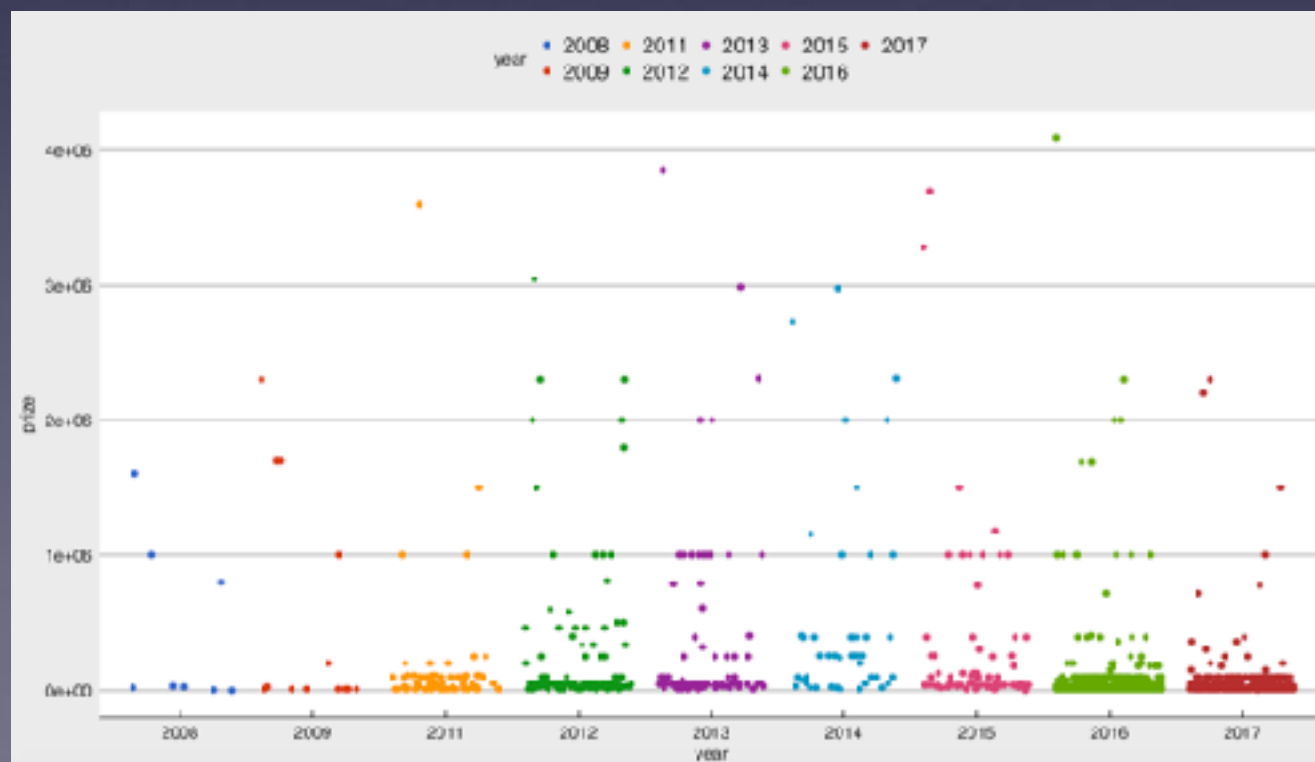
Top locations that sell the most amount of winning tickets over 10k
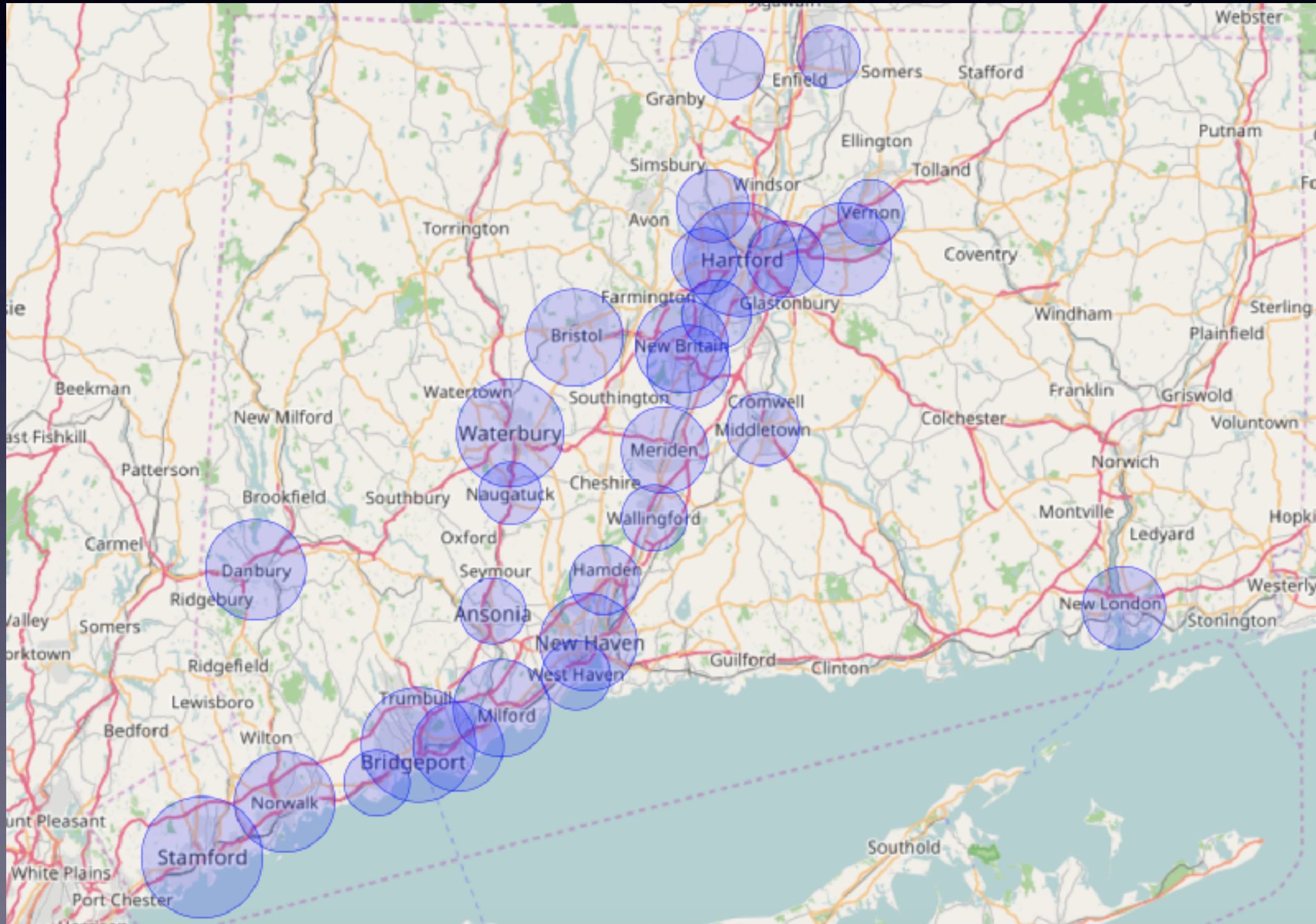
	retailer_location	count	lon	lat
1	Stamford, CT	74	-73.53873	41.05343
2	Bridgeport, CT	65	-73.19518	41.18655
3	Hartford, CT	63	-72.68509	41.76371
4	Waterbury, CT	54	-73.05150	41.55815
5	Danbury, CT	49	-73.45401	41.39482
6	Norwalk, CT	49	-73.40816	41.11774
7	Milford, CT	47	-73.06404	41.23070
8	Bristol, 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	-72.76871	41.63538
16	East Hartford, CT	28	-72.61283	41.76342
17	Middletown, CT	26	-72.65065	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

Population per sq. mile
- 1...10
- 10...25
- 25...50
- 50...100
- 100...250
- 250...500
- 500...1000
- 1000...2500
- 2500...5000
- >5000
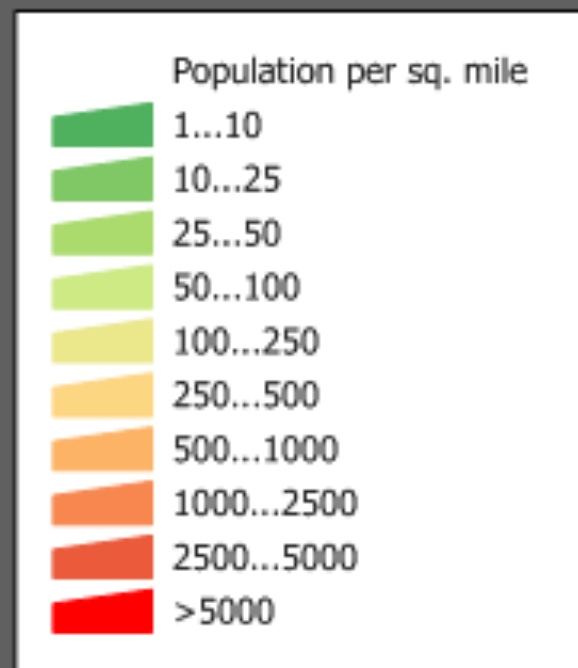
Source: U.S. Census Bureau
Census 2010 Summary File 1
population by census tract

Games

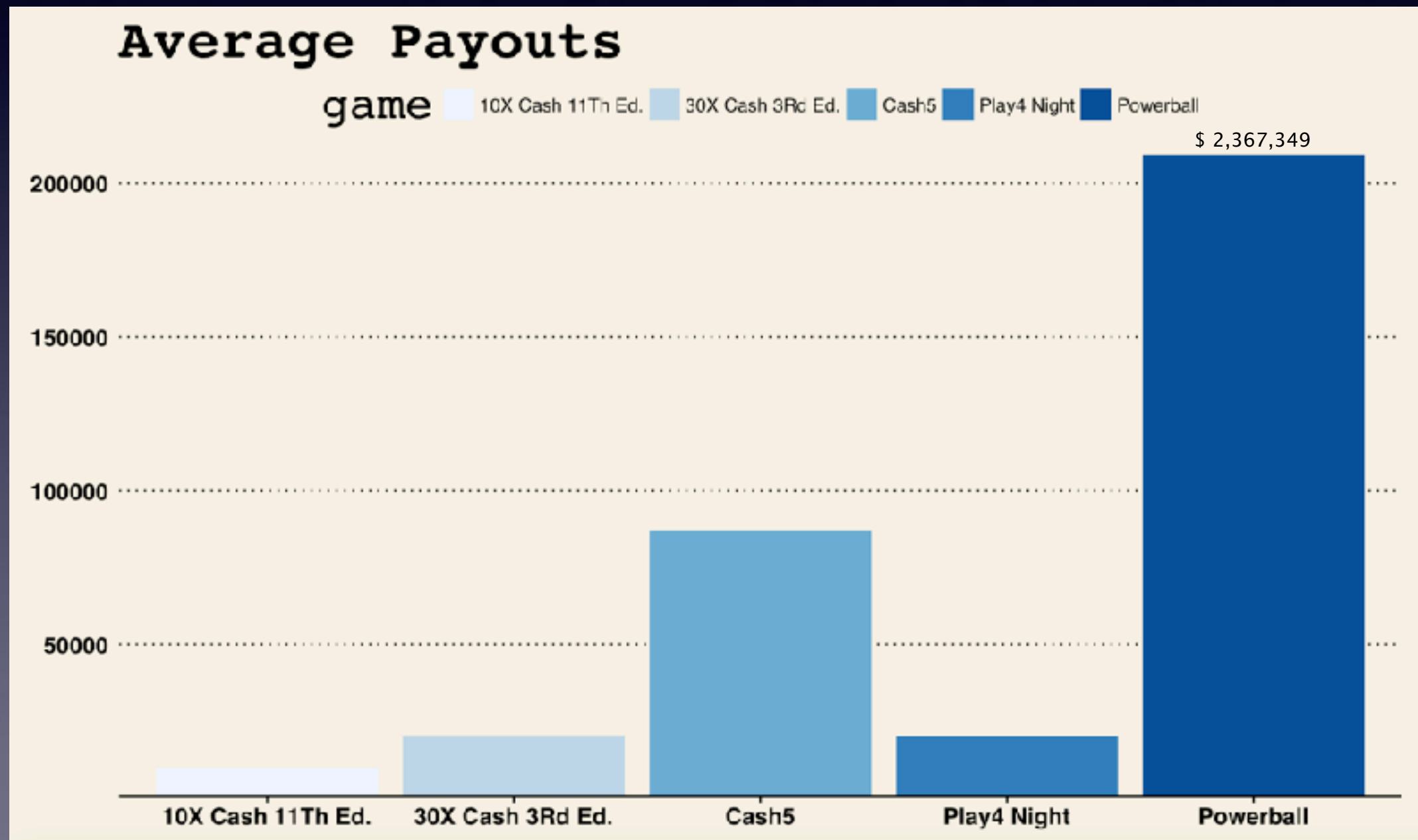Games that produce the most frequent Winners (Above 10k)
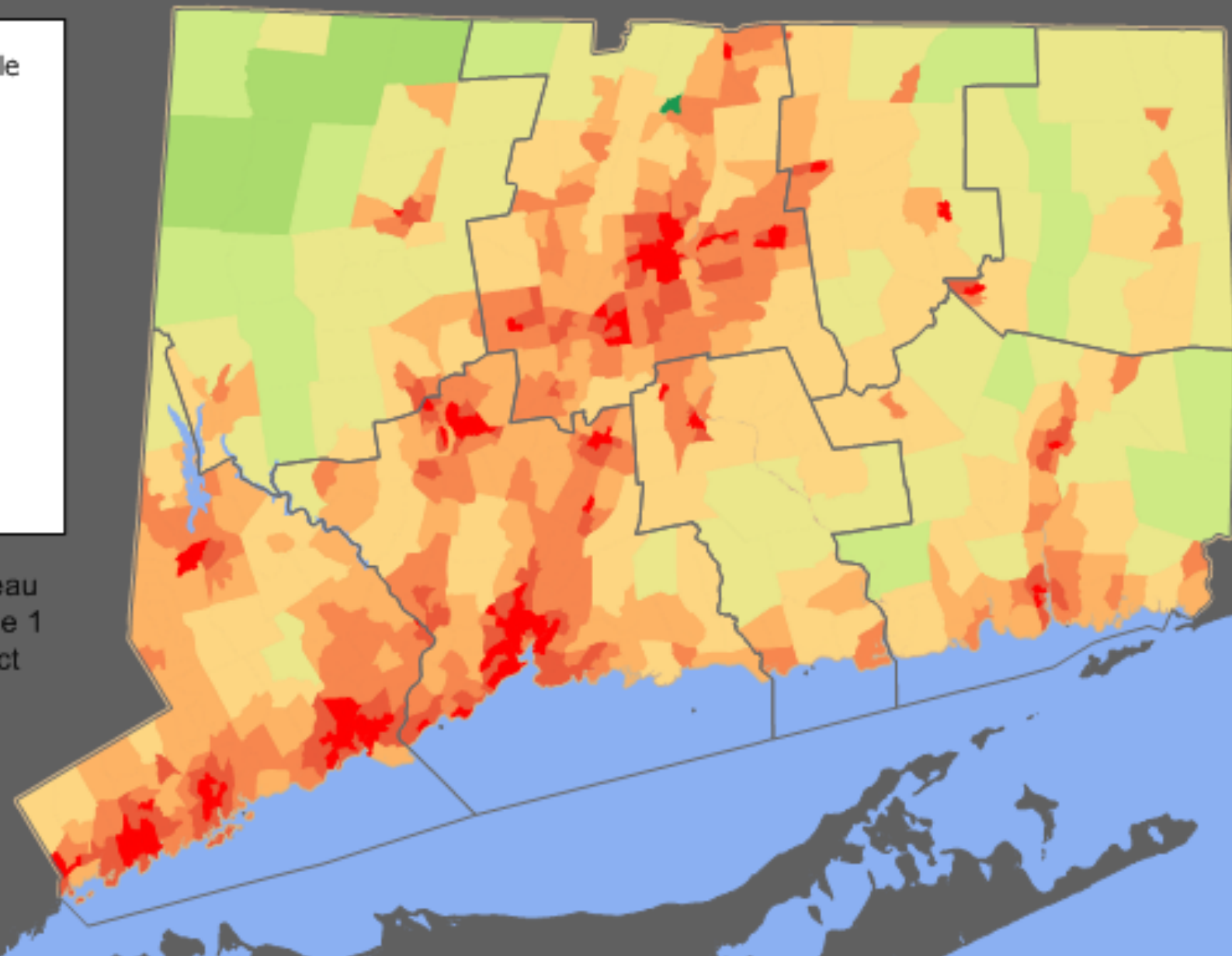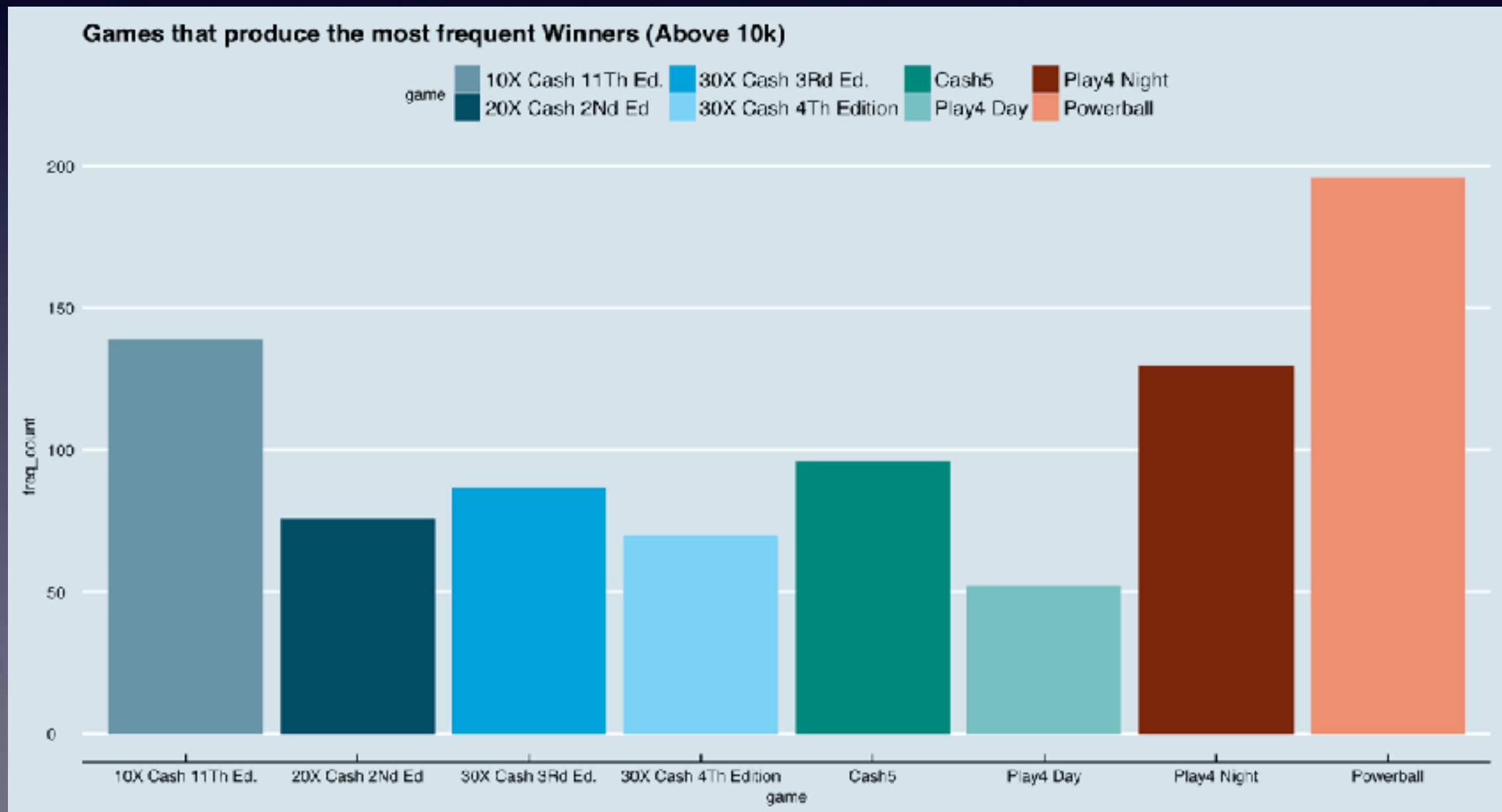
Games that pay
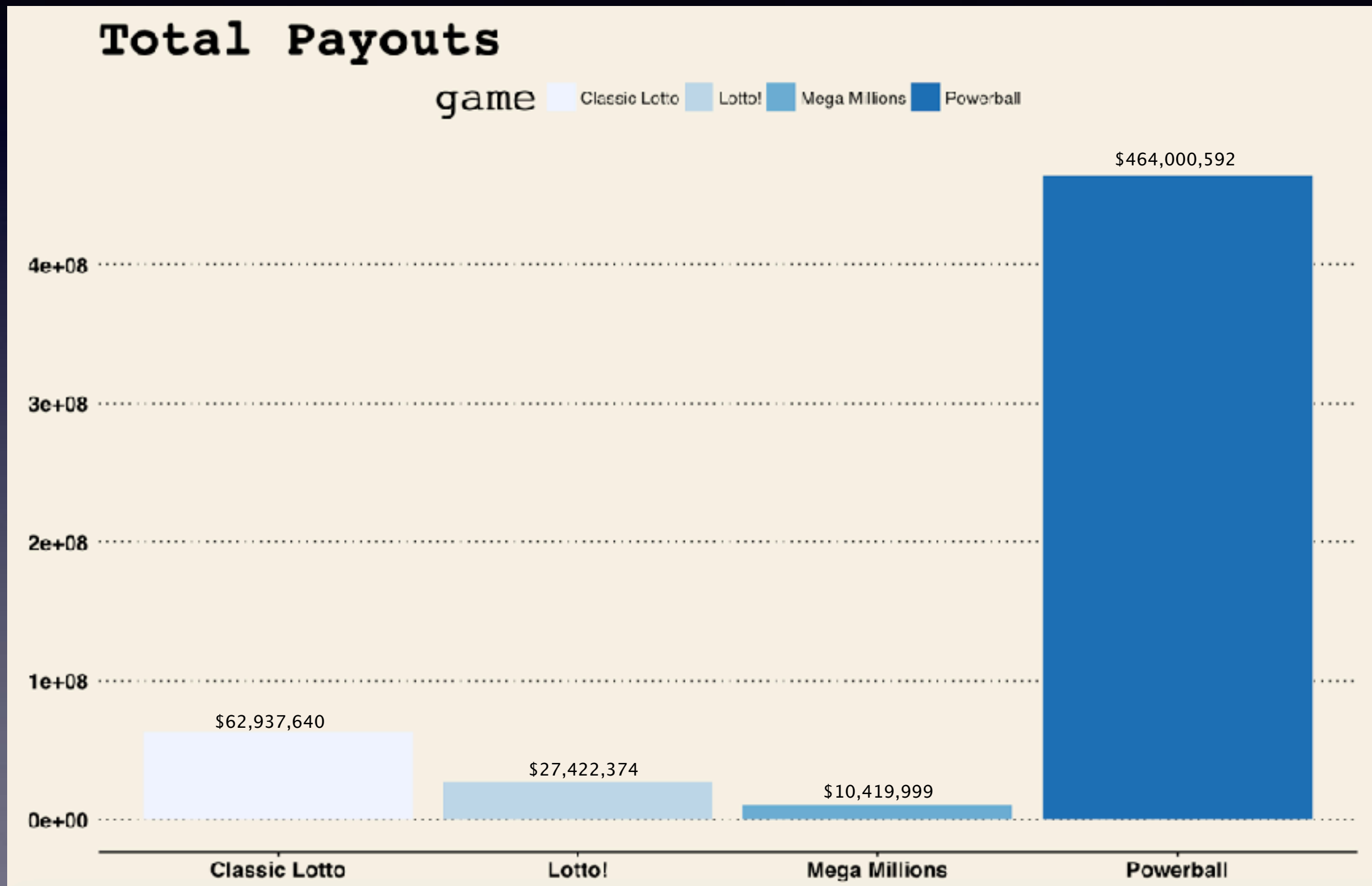
Average Payouts
for top games
(2008-Present)

Winners by Month

Monthly Number of Wins for Top Games