Approximate Computing: An effective likelihood-free method with statistical guarantees

Ryan Gross

Mentors: Dr. Minge Xie and Suzanne Thornton

DIMACS REU 2018

Project Overview

The theory: Approximate computing

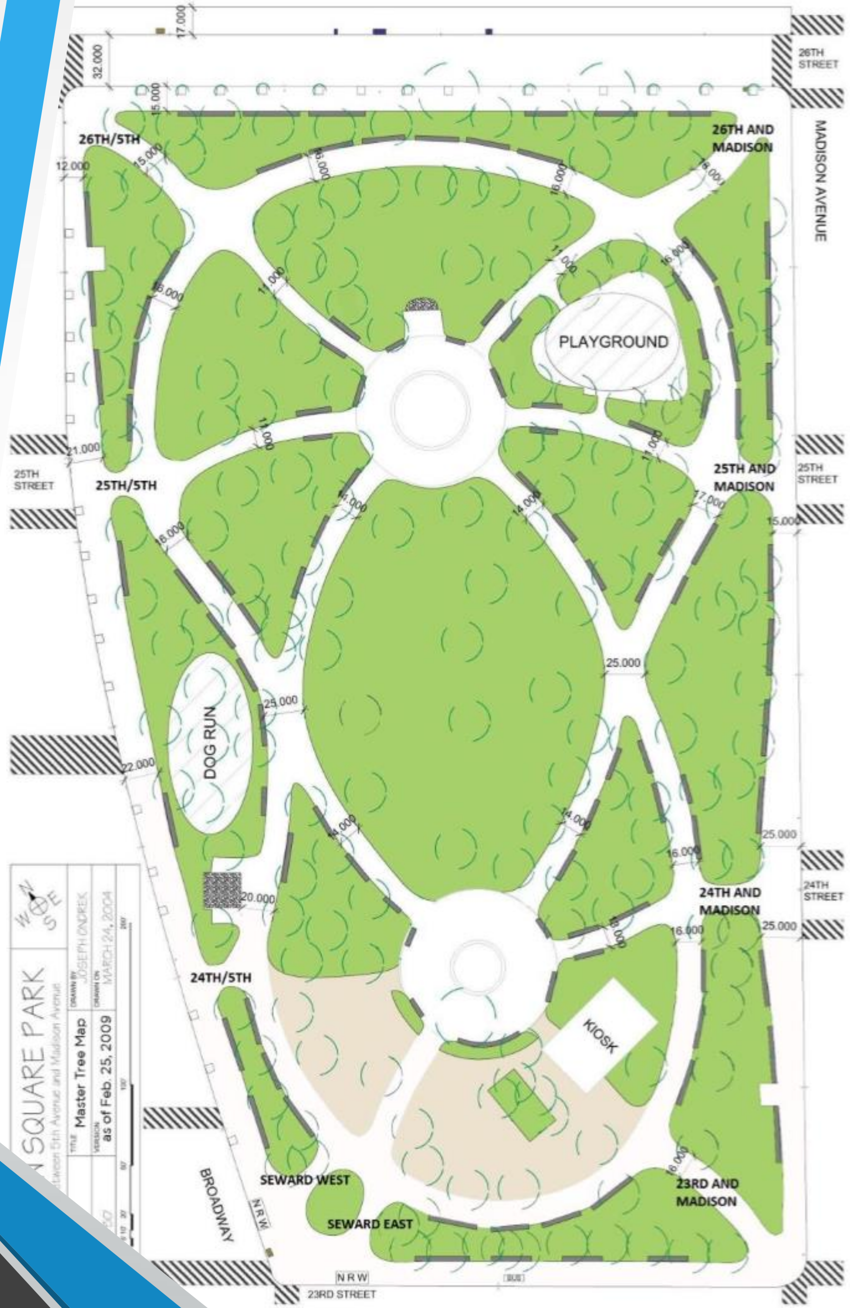
- Likelihood-free method of making statistical estimates
  - Likelihood-> probabilities associated with parameters of a set of observed data
  - Likelihood-free-> allows for creation of statistical models with complex data (e.g. population genetics, traffic flow)
- Approximate Bayesian Computing (ABC), Approximate Confidence Distribution Computing (ACC)

The application: Madison Square Park data set

- Given entrant and exit data, we look to utilize approximate computing methods to make inferences on pedestrian traffic over given time periods in Madison Square Park

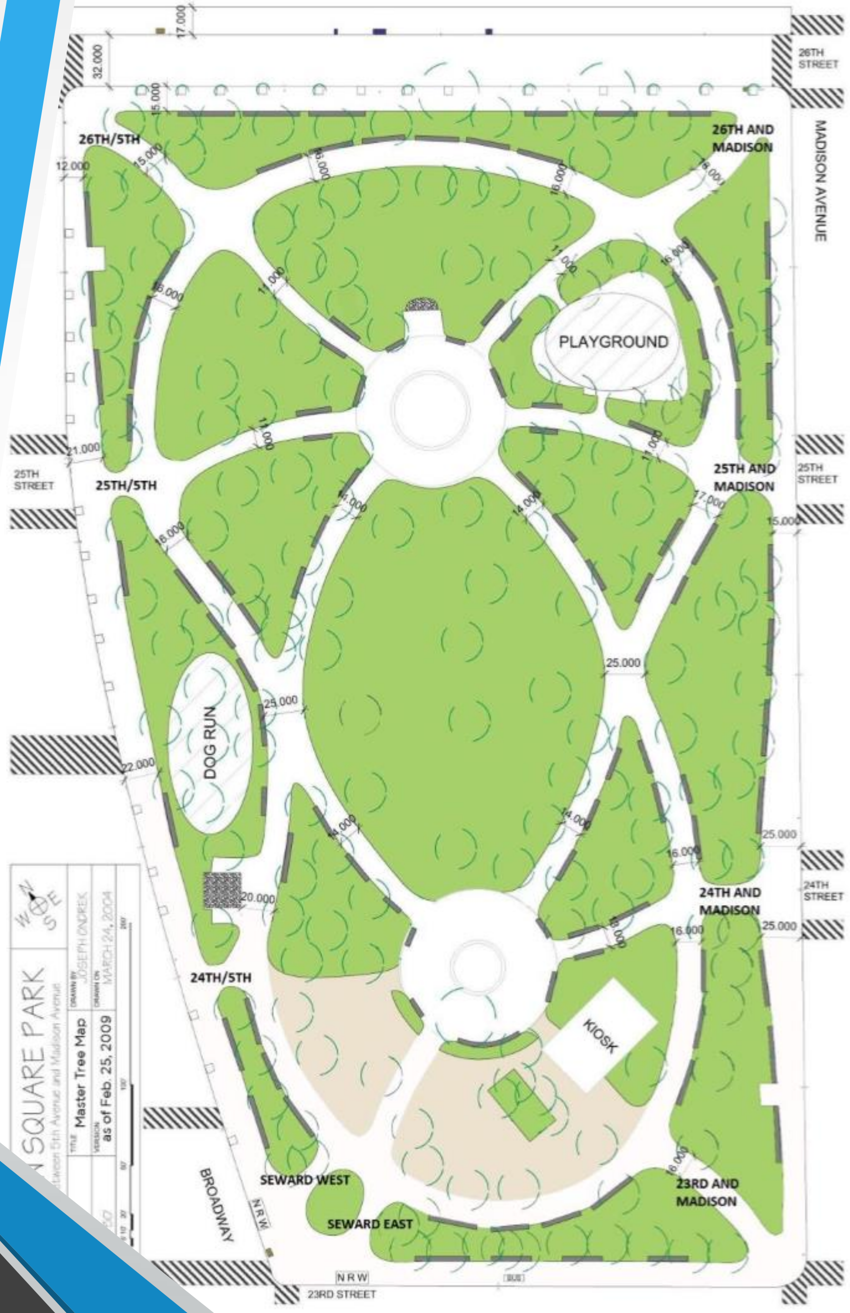
The Procedure

- Madison Square Park has 9 entrances/exits as shown
- Only 2 locations at a given time are equipped with counters to track pedestrian movement
  - Stationary counter: located at Seward East
  - Mobile counter: moves among the 8 remaining locations on a given schedule



How to estimate number of daily visitors to all 9 locations in total?

- Initial method
  - Take means of daily visitors at each location over a previous time period
  - Sum all locations to produce estimate
- Better methods
  - ABC, ACC



Challenges in the Data

- Missing data to 7 of 9 locations on a given day
- 2 “transition days” each week where mobile counter location is not constant and not recorded
- One-day events, weather, etc. skew data and introduce confounding variables
- 18 potential parameters (9 locations, total in/out) with limited number of data points
  - Can we make assumptions to limit the number of parameters?
  - Can we model/simulate traffic flow to find patterns in pedestrian activity?

179	Thu, Jun 2, 16	14666	8062	6604	Seward West
180	Fri, Jun 3, 16	11266	5957	5309	Seward West
181	Sat, Jun 4, 16	6434	3746	2688	Seward West
182	Sun, Jun 5, 16	4567	2613	1954	Seward West
183	Mon, Jun 6, 16	4730	3449	1281	Transiton
184	Tue, Jun 7, 16	1	1	0	24th/5th
185	Wed, Jun 8, 16	10	8	2	24th/5th
186	Thu, Jun 9, 16	32	15	17	24th/5th
187	Fri, Jun 10, 16	2	2	0	24th/5th
188	Sat, Jun 11, 16	7096	3637	3459	Transiton
189	Sun, Jun 12, 16	8770	4678	4092	24th/5th
190	Mon, Jun 13, 16	8881	4386	4495	24th/5th
191	Tue, Jun 14, 16	9848	4919	4929	Transiton
192	Wed, Jun 15, 16	8922	4137	4785	25th/5th

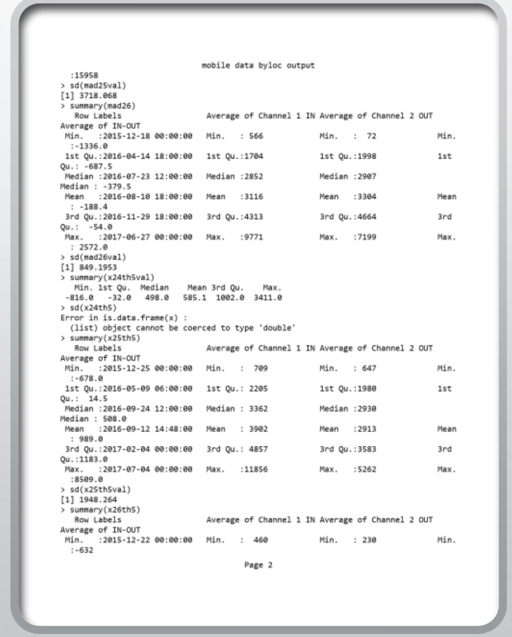
Wed, Jul 5, 17	6396	4510	1886	Transition
Thu, Jul 6, 17	7346	4693	2653	24th/5th
Fri, Jul 7, 17	5385	3554	1831	
Sat, Jul 8, 17	4168	2842	1326	
Sun, Jul 9, 17	3692	2458	1234	
Mon, Jul 10, 17	7123	4984	2139	
Tue, Jul 11, 17	7400	5195	2205	
Wed, Jul 12, 17	5654	4126	1528	
Thu, Jul 13, 17	5113	3719	1394	
Fri, Jul 14, 17	5048	2857	2191	
Sat, Jul 15, 17	4507	3060	1447	
Sun, Jul 16, 17	4048	2866	1182	
Mon, Jul 17, 17	6480	4503	1977	Counter was not consistently moved from 7/7 on
Tue, Jul 18, 17	6267	4388	1879	
Wed, Jul 19, 17	5651	4272	1379	
Thu, Jul 20, 17	5274	3951	1323	
Fri, Jul 21, 17	5632	4210	1422	
Sat, Jul 22, 17	3062	2290	772	
Sun, Jul 23, 17	3239	2062	1177	
Mon, Jul 24, 17	6265	3767	2498	
Tue, Jul 25, 17	8371	4972	3399	
Wed, Jul 26, 17	7417	3910	3507	
Thu, Jul 27, 17	9678	4585	5093	
Fri, Jul 28, 17	7608	3904	3704	
Sat, Jul 29, 17	5960	3871	2089	
Sun, Jul 30, 17	4844	3103	1741	
Mon, Jul 31, 17	8608	4684	3924	
Tue, Aug 1, 17	6041	3152	2889	
Wed, Aug 2, 17	6088	3012	3076	
Thu, Aug 3, 17	4731	2836	1895	

Research Process

Early Progress

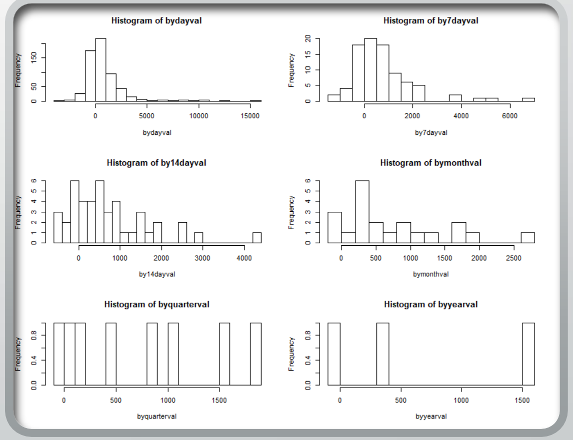
- Working in Excel and R to clean/ examine the data set
- Looking for relationships
  - Over time
  - By location
- Determining proper summary statistics
- Researching potential traffic flow models

Row Labels	Average of Total	Average of Channel 1 IN	Average of Channel 2 OUT	Average of IN-OUT
23rd/Madison	3909	2137	1772	365
24th/5th	7537	4061	3476	585
24th/Madison	9000	4574	4426	147
25th/5th	6816	3902	2913	989
25th/Madison	10696	6127	4569	1558
26th/5th	5540	3365	2175	1191
26th/Madison	6420	3116	3304	-188
Seward West	9088	4868	4376	492
Transition	9165	4980	4185	794
(blank)	7087	4319	2768	1550
Grand Total	**7670**	**4214**	**3467**	**748**



```
mobile data byloc output
  :15958
> sd(mad25val)
[1] 3718.068
> summary(mad26)
   Row Labels                 Average of Channel 1 IN Average of Channel 2 OUT
Average of IN-OUT
 Min.   :2015-12-18 00:00:00   Min.   : 566            Min.   :  72            Min.
  :-1336.0
 1st Qu.:2016-04-14 18:00:00   1st Qu.:1704            1st Qu.:1998            1st
Qu.: -687.5
 Median :2016-07-23 12:00:00   Median :2852            Median :2907
Median : -379.5
 Mean   :2016-08-10 18:00:00   Mean   :3116            Mean   :3304            Mean
  : -188.4
 3rd Qu.:2016-11-29 18:00:00   3rd Qu.:4313            3rd Qu.:4664            3rd
Qu.:  -54.0
 Max.   :2017-06-27 00:00:00   Max.   :9771            Max.   :7199            Max.
  : 2572.0
> sd(mad26val)
[1] 849.1953
> summary(x24th5val)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -816.0   -32.0   498.0   585.1  1002.0  3411.0
> sd(x24th5)
Error in is.data.frame(x) :
  (list) object cannot be coerced to type 'double'
> summary(x25th5)
   Row Labels                 Average of Channel 1 IN Average of Channel 2 OUT
Average of IN-OUT
 Min.   :2015-12-25 00:00:00   Min.   :  709           Min.   : 647            Min.
  :-678.0
 1st Qu.:2016-05-09 06:00:00   1st Qu.: 2205           1st Qu.:1980            1st
Qu.:  14.5
 Median :2016-09-24 12:00:00   Median : 3362           Median :2930
Median : 508.0
 Mean   :2016-09-12 14:48:00   Mean   : 3902           Mean   :2913            Mean
  : 989.0
 3rd Qu.:2017-02-04 00:00:00   3rd Qu.: 4857           3rd Qu.:3583            3rd
Qu.:1183.0
 Max.   :2017-07-04 00:00:00   Max.   :11856           Max.   :5262            Max.
  :8509.0
> sd(x25th5val)
[1] 1948.264
> summary(x26th5)
   Row Labels                 Average of Channel 1 IN Average of Channel 2 OUT
Average of IN-OUT
 Min.   :2015-12-22 00:00:00   Min.   :  460           Min.   : 230            Min.
  :-632

Page 2
```



References

Thornton, S., & Xie, M. (2017). Approximate confidence distribution computing: An effective likelihood-free method with statistical guarantees. arXiv:1705.10347