Suspicious Behavior at the Lekagul Preserve

VASTly Hyperbolical
Joe Hilleary, Nirmiti Naik, Gabriel Sessions

Motivation & Goals

We wanted to explore the dataset and identify any vehicle behaviors that stood out. This felt like a good starting point for investigating disruptive behavior (assumption since bird nesting was disturbed) which lead us to consider...

Analysis Questions

- What are unusual itineraries?
- Is there a reasonable explanation for the itineraries we've identified?
- How does this impact the birds?

Itinerarizing Data

car-id	car-type	time	day	vec_time	vec_date	entrance3	general-gate1	ranger-stop2	ranger-stop0	general-gate2	...	gate6	gate7	ranger-stop7	camping4	
20150001010009-284	3	13:00:09	2015-07-01	46809.0	61	46809.0	48053.0	48212.0	48396.0	48591.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-8
20150001050042-811	1	17:00:42	2015-09-01	61242.0	123	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-8
20150001060025-421	2	06:00:25	2015-09-01	21625.0	123	-86400.0	23469.0	23310.0	23126.0	22931.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-8
20150001070000-174	1	07:00:00	2015-09-01	25200.0	123	-86400.0	26799.0	26654.0	26485.0	26307.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-8
20150001070006-104	1	07:00:06	2015-08-01	25206.0	92	-86400.0	26898.0	26744.0	26566.0	26377.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-8

car-id	car-type	vec_time	vec_date	entrance3	general-gate1	ranger-stop2	ranger-stop0	general-gate2	general-gate5	entrance4	...	gate6	gate7	ranger-stop7	campir
20150001010009-284	3	53403.222222	61.222222	46809.0	48053.0	48212.0	48396.0	48591.0	49288.0	-86400.0	...	-86400.0	-86400.0	-86400.0	-8640
20150001050042-811	1	63000.166667	125.000000	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	64427.0	...	-86400.0	-86400.0	-86400.0	-8640
20150001060025-421	2	22893.714286	123.000000	-86400.0	23469.0	23310.0	23126.0	22931.0	22232.0	21625.0	...	-86400.0	-86400.0	-86400.0	-8640
20150001070000-174	1	26900.375000	123.000000	-86400.0	26799.0	26654.0	26485.0	26307.0	-86400.0	-86400.0	...	-86400.0	-86400.0	-86400.0	-8640
20150001070006-104	1	39954.636364	93.363636	-86400.0	26898.0	26744.0	26566.0	26377.0	-86400.0	-86400.0	...	-86400.0	-86400.0	-86400.0	-8640

Pre-computation (DBSCAN)

```
clusters = DBSCAN(eps=140000,min_samples=50).fit(modeling_data)
```

```
pd.Series(clusters.labels_).unique()
```

```
array([ 0,  1,  2,  3, -1,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15,
       16, 17, 18], dtype=int64)
```

- Identifies clusters
  - “Normal behavior”
- Leaves some data unclustered
  - Potential “unusual behavior”

Pre-computation (TSNE)

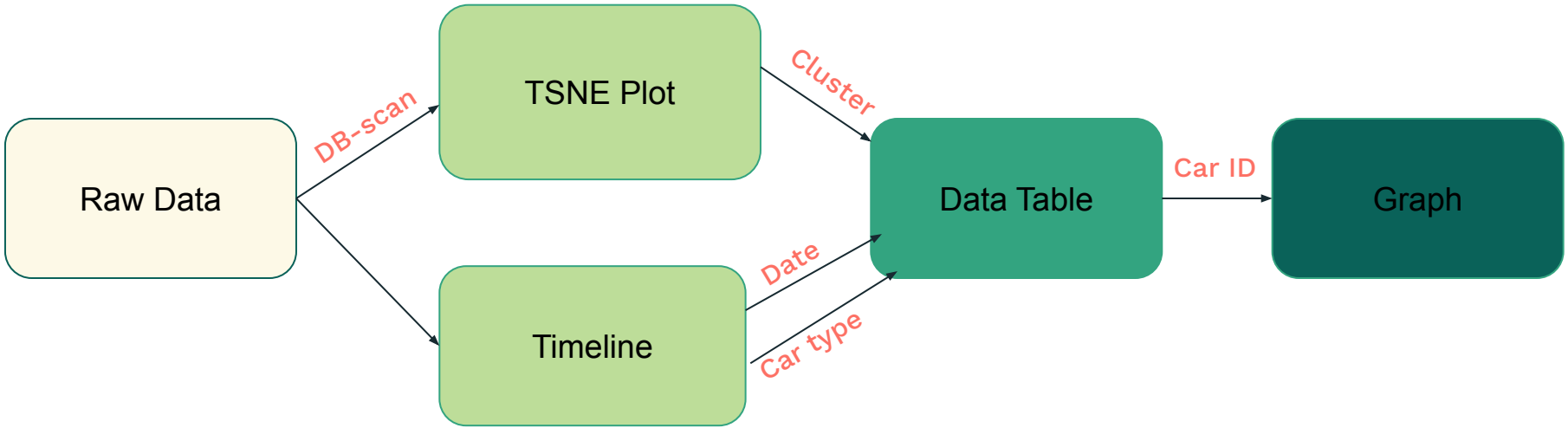
```
tsne_result = TSNE(n_iter=2000, init='pca').fit_transform(modeling_data)
```

general-gate1	ranger-stop2	ranger-stop0	general-gate2	general-gate5	...	ranger-stop7	camping4	gate2	ranger-stop1	general-gate0	camping1	gate1	gate0	tsne-2d-one	tsne-2d-two
48053.0	48212.0	48396.0	48591.0	49288.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	110985.218750	-115429.804688
-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-142643.250000	-113353.960938
23469.0	23310.0	23126.0	22931.0	22232.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	50089.847656	-117330.554688
26799.0	26654.0	26485.0	26307.0	-86400.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	42899.523438	100914.554688
26898.0	26744.0	26566.0	26377.0	-86400.0	...	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	-86400.0	42672.816406	147688.875000

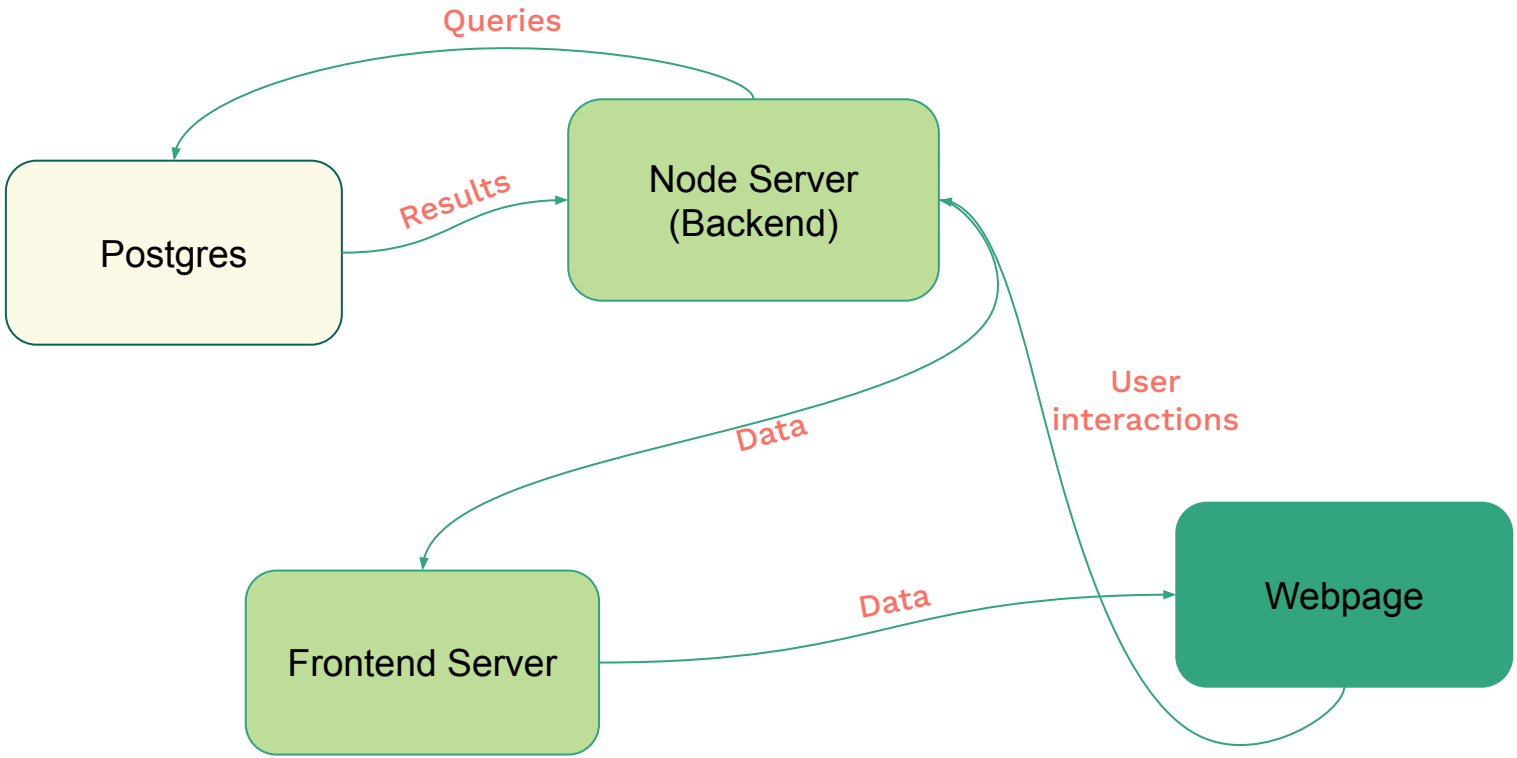
Pre-computation (Graph)

```
data = pd.read_csv('ordered_cars_id.csv')
gateNames = data.gatename.unique()

gateDict = {}
for i in range(len(gateNames)):
    gateDict[gateNames[i]] = i

numGates = len(gateNames)
adjacencyMatrix = np.zeros((numGates, numGates))

curCarID = data.head(1)["gatename"].get(0)
prev = None
for index, row in data.iterrows():
    if (curCarID != row["carid"]):
        curCarID = row["carid"]
    else:
        adjacencyMatrix[gateDict[prev["gatename"]]][gateDict[row["gatename"]]] += 1
    prev = row

print("Done Processing")

f = open("car_paths.txt", 'w')
maxValue = int(adjacencyMatrix.max())
print(maxValue)
for i in range(len(adjacencyMatrix)):
    for j in range(len(adjacencyMatrix)):
        if (int(adjacencyMatrix[i][j]) > 0):
            #f.write(str(int(adjacencyMatrix[i][j])) + " people travelled between " + gateNames[i] + " and " + gateNames[j])
            f.write("{ id: '" + str(i) + "-" + str(j) + "', source: '" + gateNames[i] + "', target: '" + gateNames[j] + "', animated: true
            f.write("\n")

f.close()
```

Data Flow

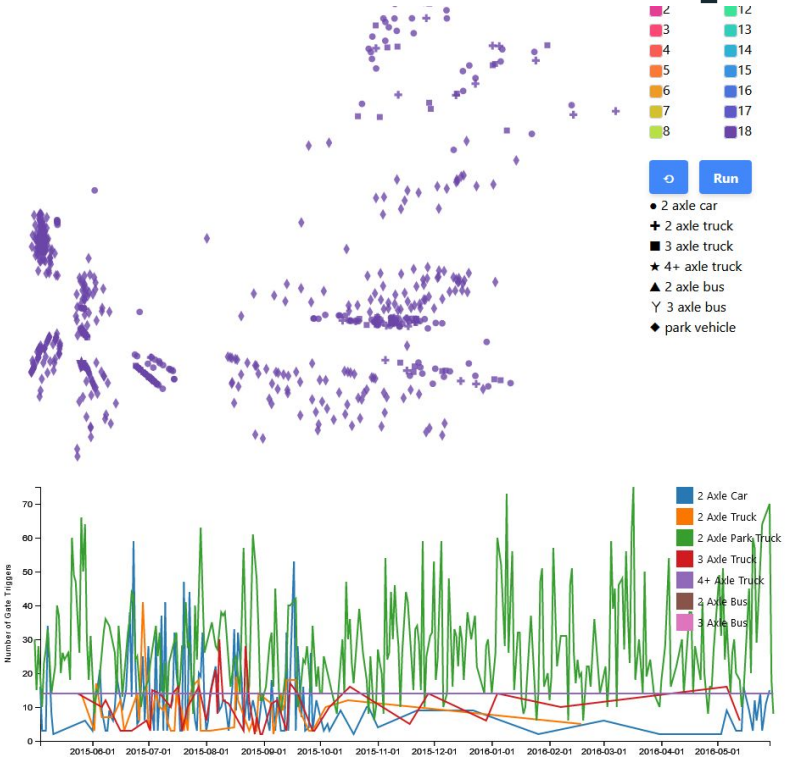

Raw Data
DB-scan
TSNE Plot
Timeline
Cluster
Date
Car type
Data Table
Car ID
Graph

Architecture

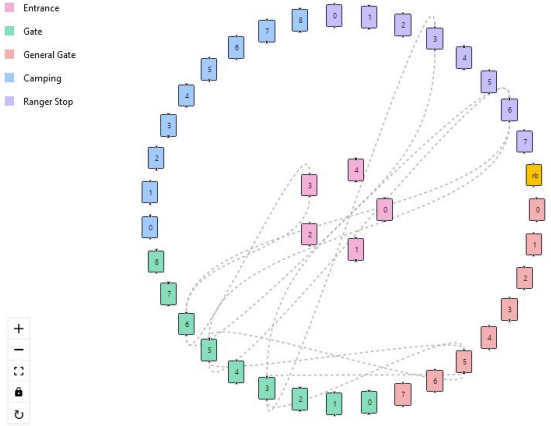


Live Demo!!!

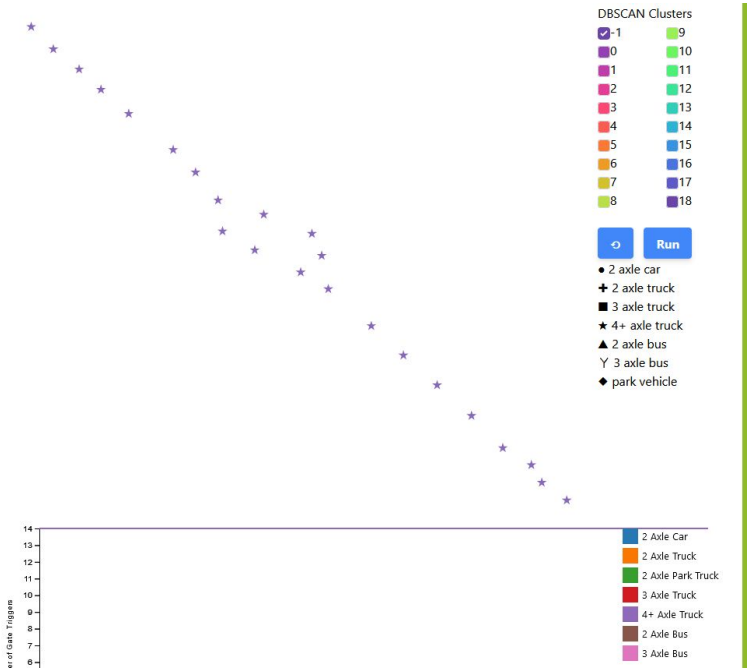
Findings: unclustered points



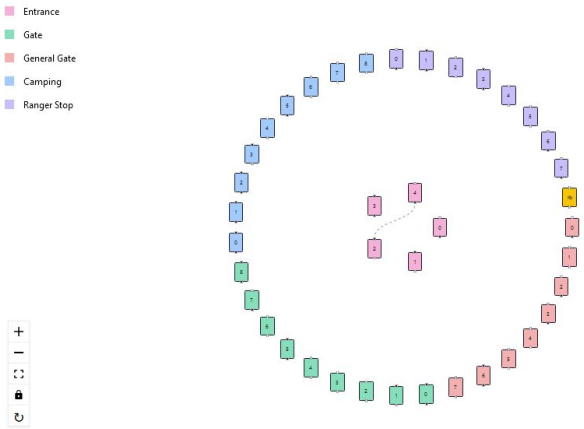
Pattern #1

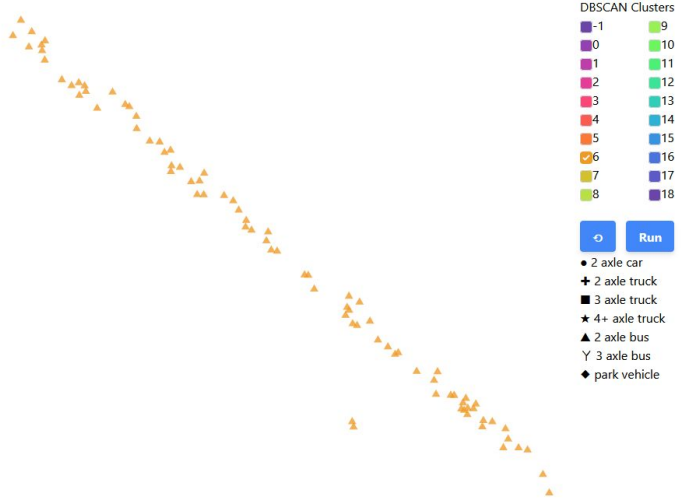


carid	cartype	cluster	first_entry	last_exit
20150104020118-228	4	-1	08/04/15 02:01:18 AM	08/04/15 02:47:00 AM
20150416040441-902	4	-1	06/16/15 04:04:41 AM	06/16/15 04:50:42 AM
20150505020522-625	4	-1	05/05/15 02:05:22 AM	05/05/15 02:57:55 AM
20150920030917-854	4	-1	10/20/15 03:09:17 AM	10/20/15 03:59:40 AM

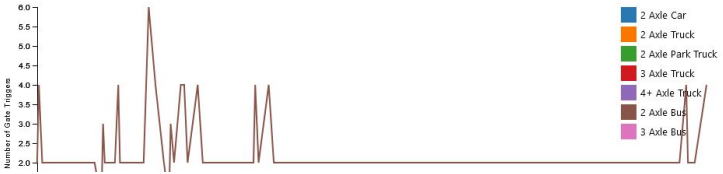


Pattern #2

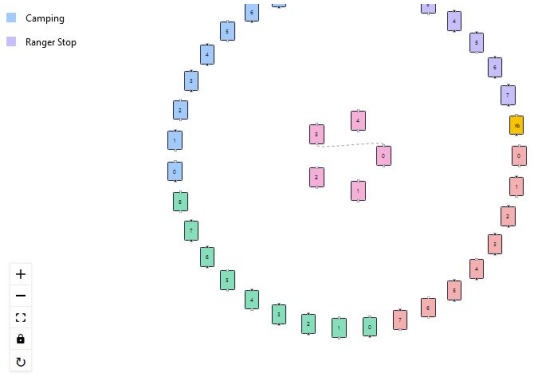


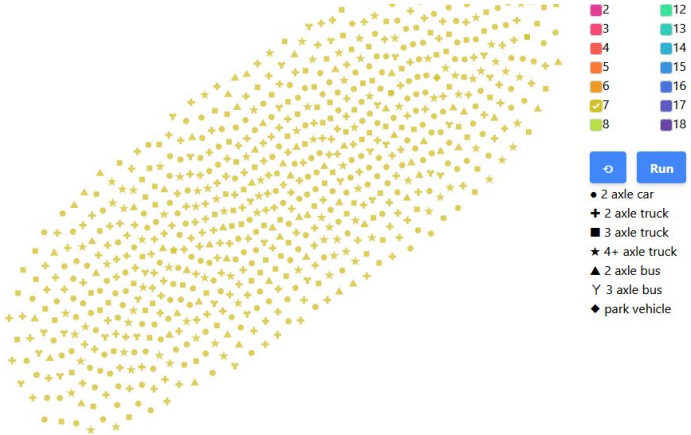


carid	cartype	cluster	first_entry	last_exit
20150004030033-995	5	6	06/04/15 03:00:33 PM	06/04/15 03:22:48 PM
20150006030042-85	5	6	09/06/15 03:00:42 PM	09/06/15 03:23:04 PM
20150225050220-518	5	6	07/25/15 05:02:20 AM	07/25/15 05:24:48 AM
20150621070647-455	5	6	11/21/15 07:06:47 PM	11/21/15 07:25:47 PM
20150627070604-127	5	6	10/27/15 07:06:04 AM	10/27/15 07:23:16 AM

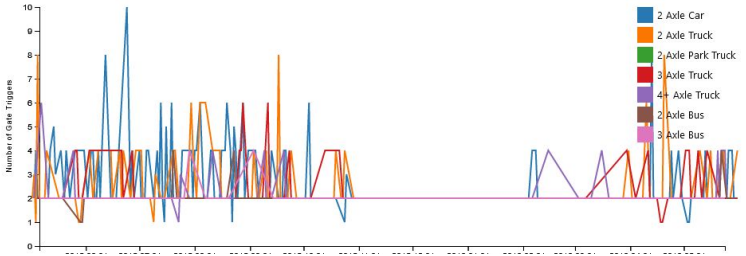


Pattern #3





carid	cartype	cluster	first_entry	last_exit
20150004050003-330	1	7	08/04/15 05:00:03 AM	08/04/15 05:25:42 AM
20150010030024-367	4	7	09/10/15 03:00:24 PM	09/10/15 03:21:03 PM
20150011060014-975	3	7	09/11/15 06:00:14 PM	09/11/15 06:21:41 PM
20150026120014-798	3	7	09/26/15 12:00:14 AM	09/26/15 12:21:14 AM
20150107060117-495	2	7	08/07/15 06:01:17 PM	08/07/15 06:21:29 PM
20150110050107-785	1	7	05/10/15 05:01:07 PM	05/10/15 05:22:04 PM
20150111090154-956	3	7	09/11/15 09:01:54 PM	09/11/15 09:22:11 PM



Discussion

- The behaviors most likely impact birds are those the occur with some consistency so that they could have a systemic impact.
- Unusual at Lekagul Reserve
  - Late night visits by large trucks (year round)
- Usual at Lekagul
  - Buses cutting through the park in the early morning (seasonal) from entrance 2 to 4
  - Regular traffic cutting through the park from entrance 3 to 0

Conclusion

The impact on bird life was likely caused by:

- Consistent vehicle activity in the middle of the night
  - Disruption of nesting and resting
- Trips of short duration
  - Disruptions in quick succession of each other

A VAST

Thank you :)

Video demonstration

Processbook