# VAST Challenge 2015 Data Description

## Background

This year’s challenge scenario is set in an amusement park similar to Hershey Park, King’s Dominion, or Busch Gardens Williamsburg. The simulated park covers a large geographic space (approx. 500x500 m^2) and is populated with ride attractions, restaurants and food stops, souvenir and game stores, an arcade, a show hall, and a performance stage. The attractions are categorized into Thrill Rides, Kiddie Rides, Rides for Everyone, Food, Restrooms, Shopping, and Shows & Entertainment. Patterns can be found in the movement through the park and communications among visitors, including expected normal visit patterns and unexpected patterns. The VAST Challenge for this year is focused on exploration of these various patterns.



## Understanding the Data

The figure above right is the park layout. The attractions are numbered and coded according to type (the index to the attractions can be found in the park website). The red line indicates the pathway through the park, although dark green areas are also areas where people can move. (Attraction 30 in the middle of the map is a water rapids ride, so people can watch from the “inside” of the ride boundaries. For other rides, people are not allowed to wander inside of the ride footprint. Attraction 63 is a show stage area, so people populate this area during performances).

## Data Source: The Park App

All visitors to the park (except for very young children) use a park app to check in to the park and rides and to communicate with fellow visitors. If visitors do not have compatible phones, they are provided with loaner devices. Visitors are assigned IDs and must use the app to check into rides and some other attractions. The park is equipped with sensor beacons that record movements within the park. Sensors are sensitive within a 5m x 5m grid cell. All pathways in the park are covered by these sensors, as are the ride check in locations. Locations are not recorded while people are on rides or inside attractions (including restaurants, stores, and rest rooms). App users may send text messages to anyone within their own designated group (for example, a family could have their own group). An app user may also make “a friend” at the park where they can send and receive texts, if both persons accept friend invitations.  
  
The datasets provided for your analysis are the movement and communication data captured from the park attendees’ apps during one weekend (Friday, Saturday, and Sunday). Structurally, the datasets are not complex, but the datasets are large and there are many patterns to reveal. In addition, some contextual data is provided for your use.

### Contextual Data

The park web site provides much contextual data. A map of the park illustrates the park layout as seen from an overhead perspective. This gives the footprint of rides and paths through the park. A map index and text descriptions of the rides and attractions are provided to give you more information about the environment within which the park visitors are traveling. The web site also provides information about activities and events in the park. A biography of all-star football (soccer) player Scott Jones is included. Scott is being honored at DinoFun World.

Also included is a news article concerning an incident at the park during the weekend for which you are provided data.

### Movement Data

The park area is gridded to assist in specifying locations. One data file contains movement information around the grid, with coordinate locations. This data appears as follow:  
  
2014-6-06 08:00:08,5231584,check-in,63,99  
2014-6-06 08:00:42,93275931,movement,64,98  
  
The values are: timestamp, person-id, type of activity (either “check-in” or “movement”), X-coordinate, and Y-coordinate.  
  
People either move from grid square to grid square, or “check in” at rides, meaning they either get in line or onto the ride. So, above, a person with ID 5231584, checked into a ride located at (63,99) at 8:00:08 AM. Person 93275931 just moved to location (64,98) at 8:00:42.  
  
As a person travels through the park, their locations are recorded in this file at a one-second resolution. If there is no record during a particular second of time, then the individual has not moved out of their previous grid square. People are not tracked after they check into a ride – they will eventually appear back on the grid when the ride is over.

### Communications Data

Another data file contains communication information. The records appear as follows:  
  
2014-6-06 08:04:01,5157734,5157729,Entry Corridor  
2014-6-06 08:14:31,715543,715544,Wet Land

2014-6-08 10:56:30,1964024,external,Kiddie Land  
  
The fields for this file are: timestamp, from (the sender ID), to (the recipient ID), and location (area where communications occurred).  
  
Location can be: Entry Corridor, Kiddie Land, Tundra Land, Wet Land, or Coaster Alley (see the Park Map on the website).

So, above, person 5157734 communicated to 5157729 from the Entry Corridor at 08:04:01 AM. Person 715543 texted person 715544 from the Wet Land at 08:14:31. Person 1964024 sent an external text to someone outside of the park at 10:56:30. External text recipients are not designated in any other way.