

No	Bash script	Experiments	Details of data generated inside folder 'data'	Plots generated inside folder 'plots'
1.	SatScanRadiusmultNA.command (Climate dataset)	Local SatScan with radius multipliers 1.1-2.9 in increments of 0.1.	Subfolder: Climate satscanradiusmultdata 1. Precision and recall for various multipliers. Precisionrecallx.txt 2. Circle details NAsatscaneighborresultx.txt 3. relevantcircles.txt for topk inside region of relevance. 4. seattlesatscannormhighlowrectanglestopk.txt for rectangles inside top k circles.	Subfolder: Climatesatscannormalwithradiusmultplots satscannormalwihradiusmulttop400mult2.4.pdf Top 400 circles are plotted for radius multiplier 2.4.
2.	SatScanRadiusMultTraffic.command	Local SatScan with radius multiplier 1.1 to 2.9.	Subfolder: TrafficSatScanradiusmultdata 1. TrafficsatscaneighborresultX.txt circle details for the X multipliers.	Subfolder: Trafficsatscannormalwithradiusmultplots Trafficradiusmult2.0.pdf for top 10 circles for radius multiplier 2.0
3.	SatScansoftwareonclimate.command	Sat Scan software. a) Bernoulli on NA(cases are temp <=0.0 or >=25.0) b) Normal on NA c) Normal on rectangular region in Washington (Normal for high and low. No restriction for the clusters. Parameters can be changed by modifying SatScanNormalParser.java and arguments passed to it.)	Subfolder: ClimateSatScansoftwareinputdata 1. Inputs for the satscan software.ctl, .cas, .geo generated from tempNA.txt for NA and for Washington area. Subfolder: SatScansoftwareoutputdataBernNA 1. Output data from SatScan for Bernoulli for NA Subfolder: SatScansoftwareoutputdataNormNA 1. Outputdata from SatScan for Normal on NA 2. Precision and recall from top 10 until recall is 1.0. Precisionrecall0.txt 3. Precision recall from top 10 to 100. Precisionrecallsatscan100.txt 4. relevantcircles.txt for the one for topk for which precisionrecall called. 5. seattlesatscanhighlowrectanglestopk.txt for rectangles in topk returned areas. Subfolder: SatScansoftwareoutputdataNormSeattle The same as above but this is done for the rectangular area in Washington.	Subfolder: SatScansoftwareplotsBernNA SatScansoftware100.pdf top 100 circles Subfolder: SatScansoftwareplotsNormNA SatScansoftware100.pdf Subfolder: SatScansoftwareplotsNormSeattle SatScansoftware100.pdf
4.	SatScansoftwareontraffic.command	Sat Scan software 1) Bernoulli (cases are for speed <= 10.8 km/h) 2) Normal(For low)	Subfolder: TrafficSatScansoftwareinputdata 1. Inputs for stascan software software Subfolder: TrafficSatScansoftwareoutputdataBern Outputs from SatScan software Subfolder: TrafficSatScansoftwareoutputdataNorm Outputs from SatScan software	Subfolder: SatScansoftwareplotsTrafficBern SatScansoftwareBernTraffic.pdf Subfolder: SatScansoftwareplotsTrafficNorm SatScansoftwareNormtraffic.pdf
5.	ClimateGinircircles.command	Growing circular regions around each point by gini coefficient. Circles enlarged till the gini coefficient is 0.30.	Subfolder: ClimateGinircirclesdata 1. ablinesXs.txt, ablinesYs.txt for the grid lines. 2. giniareastrings.txt – unsorted regions 3. NAginircirclescaseX.txt – sorted by the 4 cases. 4. precisionrecallcaseX.txt for the 4 cases. 5. releveantcirclesX.txt for the 4 cases 6. groundtruthseattlelatlongrectangles.txt 7.seattleginircirclecaseXrectanglek for the top k circles of each of	Subfolder: ClimateGinircirclesplots 4 plots ClimateGinitop400caseX.pdf for the 4 cases for the top 400 circles.

			the 4 cases.	
6.	TrafficginiCircles.command	Growing circular regions around each point by gini coefficient. Circles enlarged till the gini coefficient is 0.30.	Subfolder: TrafficGiniCirclesdata 4day10hourginiCircles.txt for the circles. For each data point the largest circle with the gini index below threshold is taken.	Subfolder: TrafficGiniCirclesplots TrafficginiCircleCaseX.pdf for the 4 sorting criteria(refer paper)
7.	GiniareabestpairNA.command	Growing regions combining the best pair during each iteration.	Subfolder: ClimateginiBestPairAlgorithmNAdata 1. ablinesXs.txt, ablinesYs.txt for the grid lines. 2. giniareastrings.txt – unsorted regions 3. giniareastringsgini.txt – sorted by gini index – case 1 4. giniareastringswithneighborsavg.txt – sorted by average – case 2 5. giniareastringsabswithneighbors.txt – sorted by abs value – case 3 (refer paper) 6. giniareastringsdiff.txt – sorted by diff between gini with next – case 4 (refer paper) 7. climateginiarearectangles – rectangles of the gini areas – empty line between different gini areas. 8. precisionandrecallX.txt for the 4 cases. 9. precisionandrecallXwithlimit50.txt for top 50 gini areas for the 4 cases – this is as per the scatter plots generated. 10. groundtruthseattlelatlongrectangles.txt	Folder: ClimateginiBestPairAlgorithmNAplots 4 plots giniareathresholdtop50caseX.pdf for the 4 cases.
8.	Giniareabestpairseattle.command	Growing regions combining the best pair at for each iteration in a rectangular region in Seattle.	Subfolder: ClimateginiBestPairSeattleAlgorithmdata 1. ablinesXs.txt, ablinesYs.txt for the grid lines. 2. giniareastrings.txt – unsorted regions 3. giniareastringsgini.txt – sorted by gini index – case 1 4. giniareastringswithneighborsavg.txt – sorted by average – case 2 5. giniareastringsabswithneighbors.txt – sorted by abs value – case 3 (refer paper) 6. giniareastringsdiff.txt – sorted by diff between gini with next – case 4 (refer paper) 7. climateginiarearectangles – rectangles of the gini areas – empty line between different gini areas. 8. precisionandrecallX.txt for the 4 cases. 9. precisionandrecallXwithlimit50.txt for top 50 gini areas for the 4 cases – this is as per the scatter plots generated. 10. sortedginiCouplesOfXiter.txt for 774 iterations. 1 for the 3 rd column- the gini areas, 2 its neighbor. 2 nd column is the id of the rank of that pair. 11. groundtruthseattlelatlongrectangles.txt	Subfolder: ClimateginiBestPairSeattleAlgorithmNAplots 4 plots giniareabpairSeattletop50caseX.pdf for the top 50 areas for the 4 cases.
9.	GiniareathresholdNA.command	Growing regions combining all pairs which form a gini index below a threshold for each iteration	Subfolder: ClimateginiThresholdAlgorithmNAdata 1. ablinesXs.txt, ablinesYs.txt for the grid lines. 2. giniareastrings.txt – unsorted regions 3. giniareastringsgini.txt – sorted by gini index – case 1 4. giniareastringswithneighborsavg.txt – sorted by average – case 2 5. giniareastringsabswithneighbors.txt – sorted by abs value – case 3 (refer paper) 6. giniareastringsdiff.txt – sorted by diff between gini with next – case 4 (refer paper) 7. climateginiarearectangles – rectangles of the gini areas – empty line between different gini areas.	Subfolder: ClimateginiThresholdAlgorithmplotsNA 4 plots giniareathresholdtop50case1.pdf for the 4 cases.

			<p>8. precisionandrecallX.txt for the 4 cases.</p> <p>9. precisionandrecallXwithlimit50.txt for top 50 gini areas for the 4 cases – this is as per the scatter plots generated.</p> <p>10. sortedginicouplesofxiter.txt for 7 iterations. 1 for the 3rd column- the gini areas, 2 its neighbor. 2nd column is the id of the rank of that pair.</p> <p>11. groundtruthseattlelatlongrectangles for the rectangles of the ground truth.</p>	
10.	Trafficginibpair.command	Grow regions combining best pair at a time	<p>Subfolder: trafficginibestpair</p> <p>1. trafficbeijingablins – ablins for the grid lines.</p> <p>2. 4day10hourbeijingtrafficgridaverages.txt – the average is computed of the speeds of all vehicles inside each grid cell.</p> <p>3. 4day10hourgridginirectanglescombined.txt for the rectangles of homogeneous regions. There is an empty line between regions.</p> <p>4. giniareastringsdiff.txt – the regions sorted by case 4.(refer paper)</p>	<p>Subfolder: trafficginibestpairplots</p> <p>Trafficginiareacombinedplot.pdf</p> <p>(The light blue circles denote the magnitude of the average speed in each grid cell. Only homogeneous regions of size more than one grid cell are marked. The rectangles of a single such region use the same color.) The region with the lowest gini index is marked with dotted black line.</p>
11.	Analyzeclimatedata.command	Analyze climate data – draw scatter plot of entire data set(world) and subset in North America. Also gives histograms and other statistics	<p>Subfolder: Climateanalysisdata</p> <p>1. tempNA.txt</p> <p>2. tempWorld.txt</p> <p>CartesiaX, CartesianY, Mean annual temperature, Longitude, Latitude</p> <p>Are the attributes in each file.</p>	<p>Subfolder: Climateanalysisplots</p> <p>1. NAhistogramoftemp.pdf</p> <p>2. NAscatteroftemp.pdf</p> <p>3. Worldhistogramoftemp.pdf</p> <p>4. Worldscatteroftemp.pdf</p>
12.	TogenerateSatScaninputsforclimate.c ommand	To generate the case, control and geo files for input to SatScan software for Bernoulli and Normal model.	<p>Subfolder: ClimateSatScansoftwareinputdata</p> <p>Bernoulli for NA</p> <p>1. NAtemplatlongberctl.ctl</p> <p>2. NAtemplatlongberncas.cas</p> <p>Bernoulli for Seattle</p> <p>3. NAtemplatlongberctlseattle.ctl</p> <p>4. NAtemplatlongberncasseattle.cas</p> <p>Normal for Seattle</p> <p>5. NAtemplatlongcasnormseattle.cas</p> <p>Normal for NA</p> <p>6. NAtemplatlongcasnorm.cas</p> <p>Geo for Seattle and NA</p> <p>7. NAtemplatlonggeo.geo</p> <p>8. NAtemplatlonggeoseattle.geo</p>	
13.	TogenerateSatScaninputsfortraffic.co mmand	To generate the case, control and geo files for input for SatScan software for Bernoulli and Normal model.	<p>Subfolder: TrafficSatScansoftwareinputdata</p> <p>1. Trafficlatlongberctl.ctl</p> <p>2. Trafficlatlongberncas.cas</p> <p>3. Trafficlatlongcasnorm.cas</p> <p>4. Trafficlatlonggeo.geo</p>	
14.	Clean.command	Clean the folders data and plots. The folder data has some essential data which will not be deleted but all sub folders are removed.	<p>Delete the subfolders inside folder “data” and “plots”.</p> <p>The following data files inside data will remain and should not be deleted.</p> <p>1. ablineXs.txt</p> <p>2. ablineYs.txt</p> <p>3. cai_temp.txt</p> <p>4. groundtruthseattle.txt</p>	<p>All sub folders inside “plots” are deleted.</p> <p>The following plot should be given to subjects for collecting groundtruths and should not be deleted:</p> <p>1. climateNAofSeattle.pdf</p>

			5. NAtemplatlongcasnorm.cas 6. NAtemplatlonggeo.geo 7. tempNA.txt 8. Trafficlatlongcasnorm.cas 9. Trafficlatlonggeo.geo	
--	--	--	---	--