Oiled Oceanic Ornithological Observations

Deepwater BP Oil Spill Data Investigation

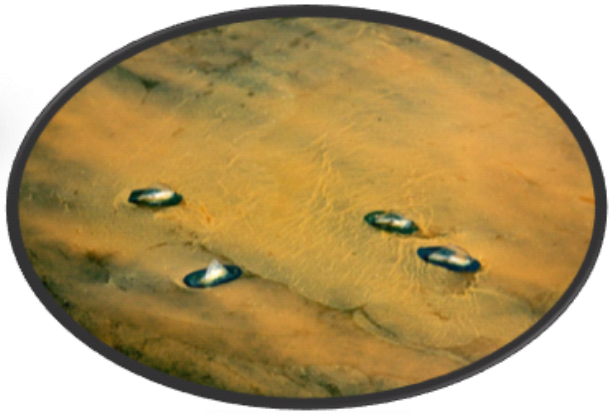
Li Wang Xiaoyue Cheng Karsten Maurer Dai-Trang Le

Department of Statistics
Iowa State University

April 27, 2011







nature news

Published online 21 April 2011 | Nature |
doi:10.1038/news.2011.250

Oil spills underreported in Gulf of Mexico

By Melissa Gaskill

"But our image analysis shows the leak rate must be greater, possibly by a factor of 10."

Original Dataset

The dataset consists of 6 variables.

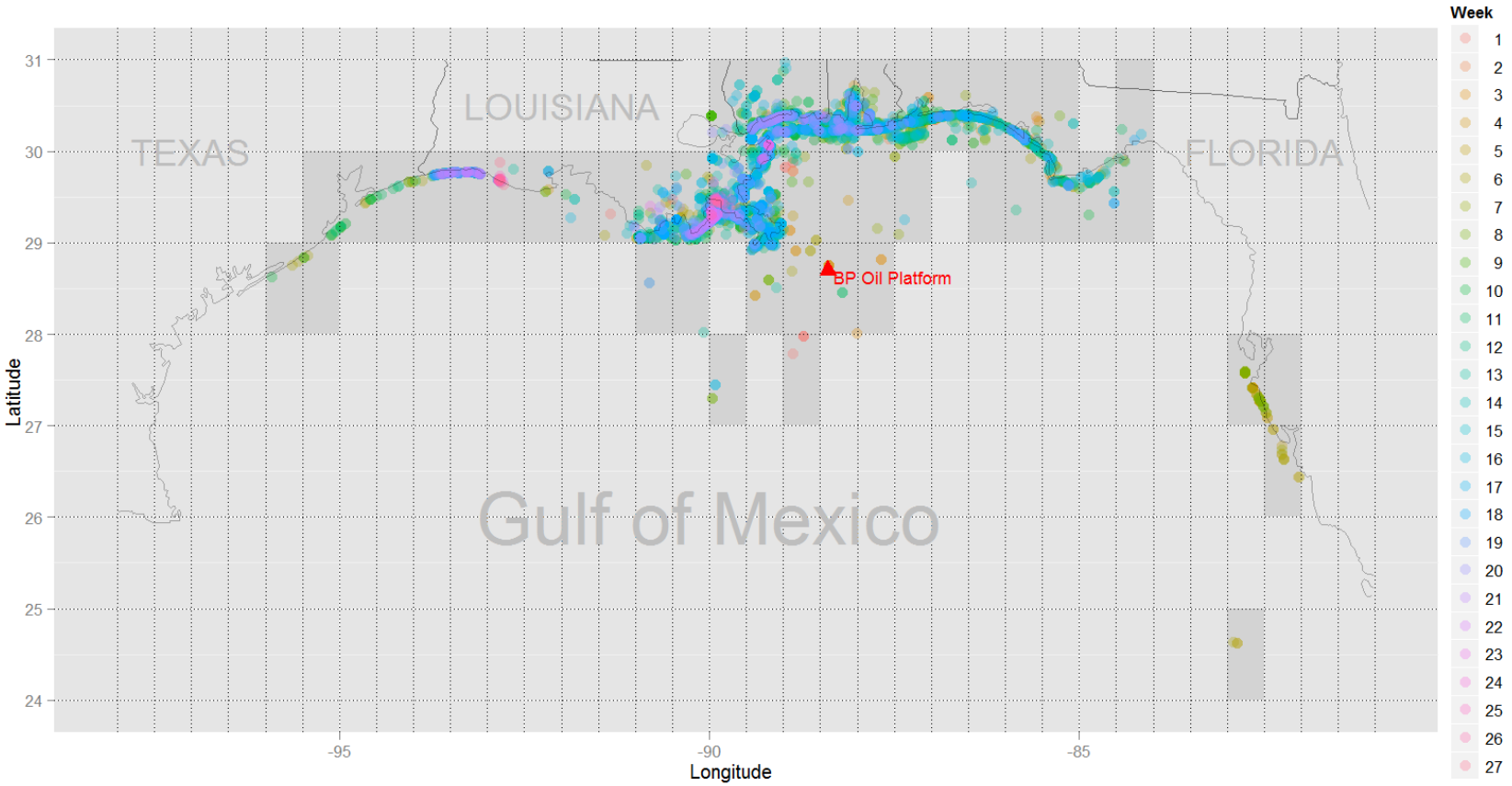
- **Species:** species of the bird found.
- **Latitude, Longitude:** location where the bird was found.
- **Oiling:** oiling condition on the bird. There are 3 levels: "Not Visibly Oiled", "Unknown", and "Visibly Oiled".
- **Condition:** the bird condition, including two levels: "Dead" and "Live".
- **Date:** the date when the bird was found, from May 5 to Nov 6.

Data Source: http://gomex.erma.noaa.gov/erma.html

Analysis Plan

- **Data Restructuring**
  - bin the locations to a few areas
  - bin the date by week
  - aggregate the counts of death and live birds
- **Summary Statistics**
- **Plots**
  - maps
  - timeplot
  - animation
- **Modeling**
  - spatial clustering

Data Aggregation



Totally 27 consecutive weeks and 47 not-all-adjoining areas

Finally obtain 435 bins $< 27 \times 47$.

Two Levels Of Data – Original Version

No.	Long	Lat	Date	Species	Oiling	Cond
1	-89.0	30.2	7/14	Gull	Unknown	Live
2	-85.6	30.3	5/17	Gannet	Unknown	Live
3	-90.0	29.3	9/11	Pelican	Unknown	Dead
...	...	...	...	...	...	...
7610	-89.3	29.9	7/19	Gull	Visibly Oiled	Dead
Summary	(Map)		Min:5/5 1stQu:7/11 Med:8/2 3rdQu:8/22 Max:11/6	Gull:3544 Pelican:877 Tern:766 Gannet:569 Other:1854	Visibly:2942 Unknown:761 Not Visibly:3907	Dead:5783 Live:1827

Two Levels Of Data – Aggregated Version

	Gridx	Gridy	Week	Count	#dead	#oiled	**DeathPct**	**OiledPct**
1	1	5	6	2	2	0	**100%**	**0%**
2	1	5	10	1	1	0	**100%**	**0%**
3	2	5	6	2	2	1	**100%**	**50%**
...	...	...	...	...	...	...	...	...
435	28	4	7	2	2	0	**100%**	**0%**
		Mean		17.5	13.3	6.8	73.8%	30.5%
		Std		33.3	27.7	17.2	29.6%	34.0%
		Min		1	0	0	0%	0%
		Median		5	4	1	80%	20%
		Max		234	226	164	100%	100%

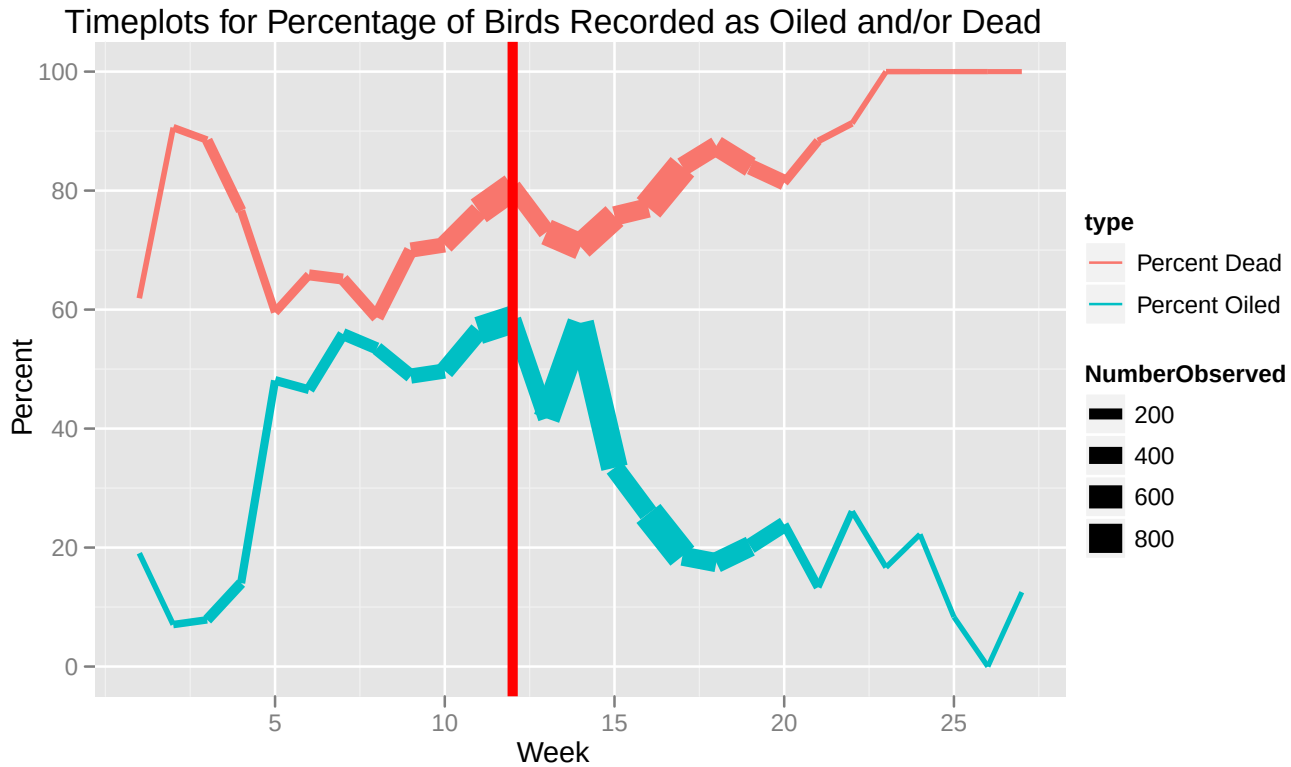
Oiling VS. Death

	Dead	Live	DeathPct %
Visibly Oiled	1984	958	67.44
Unknown	759	2	99.74
Not Visibly Oiled	3040	867	77.81
OiledPct %	34.31	52.44	

Death And Oiled Percents By Species

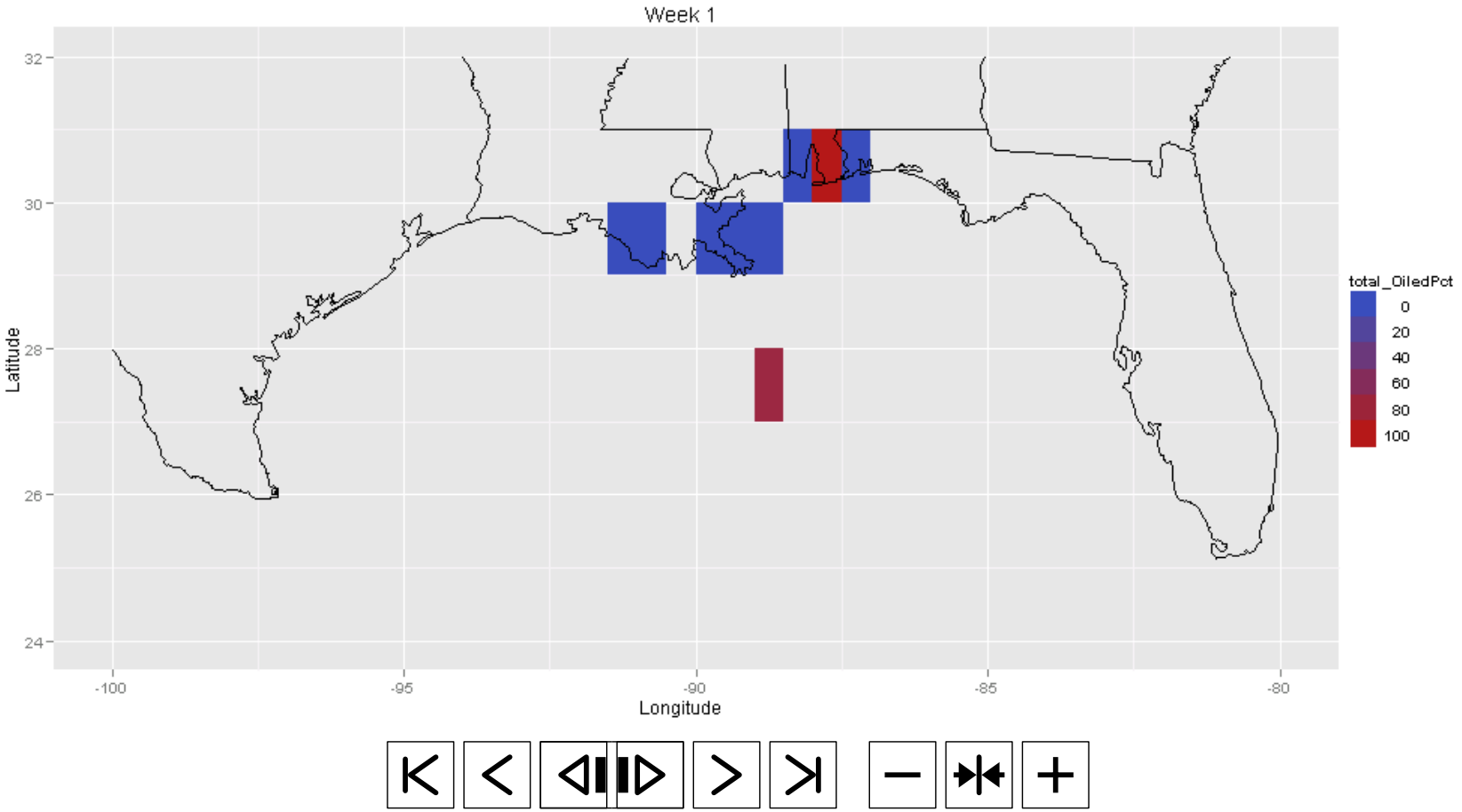
	Count	# Dead	DeathPct %	# Oiled	# NotOil	OiledPct %
Total	**7610**	**5783**	**75.99**	**2942**	**3907**	**38.66**
Gull	3544	2840	**80.14**	1379	1829	38.91
Pelican	877	516	58.84	374	378	42.65
Tern	766	580	75.72	356	343	46.48
Gannet	569	308	54.13	372	176	**65.38**
Other	1854	1539	**83.01**	461	1181	24.87

Timeplot

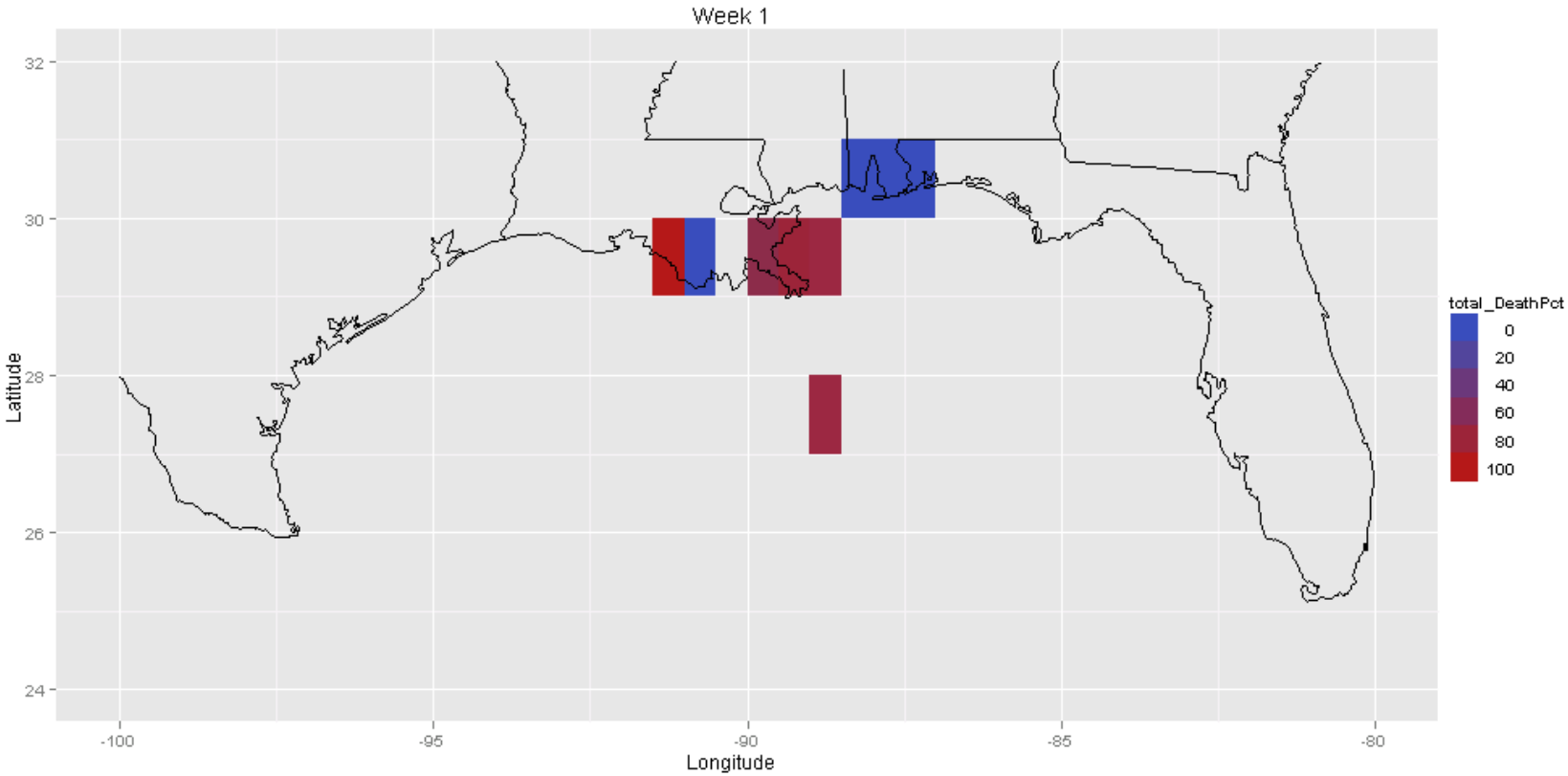


Timeplot of percentages of birds recorded as oiled and/or dead. The vertical red line indicates the time when the oil well was capped, we see the number of observations increase up until this point then begin to decrease significantly a few weeks after. The percentage oiled rises to about 60% then quickly drops after the well was capped, whereas the percent found dead spikes initially then grows consistantly.

Animated Map of Oiled Percentage In 27 Weeks

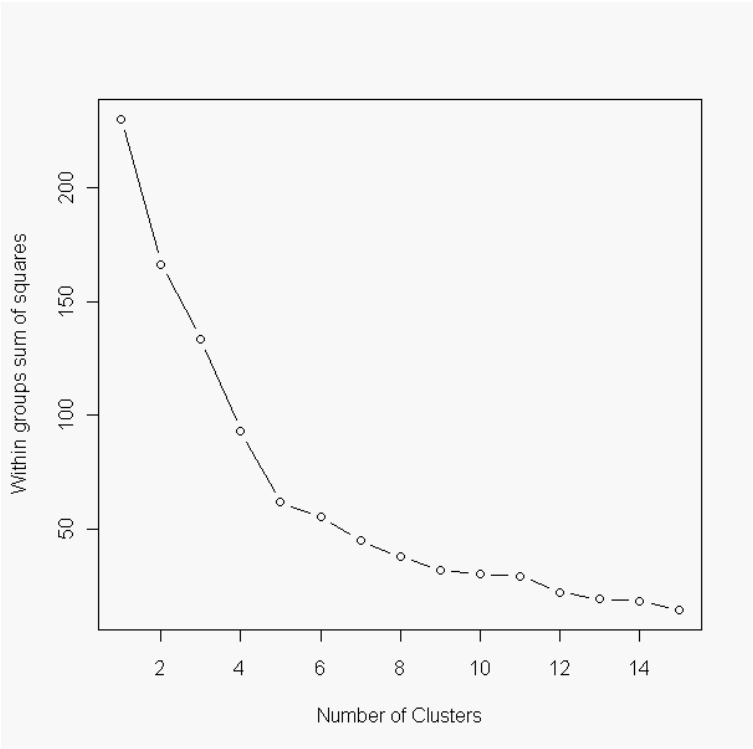


Animated Map of Death Percentage In 27 Weeks



Clustering Techniques

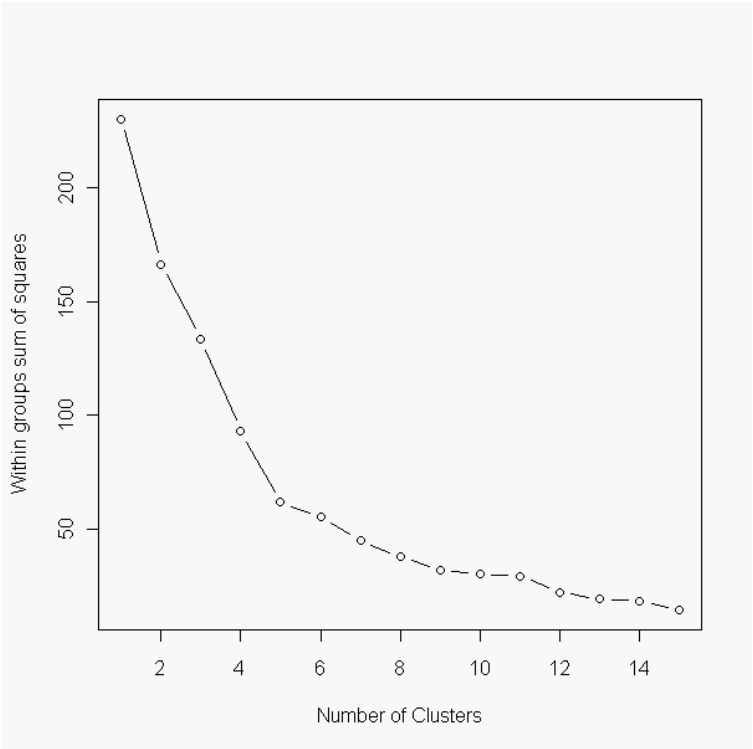
- Explore the scree plot



- A sharp bend occurs at end of the fourth segment.
- Start out with five clusters.

Clustering Techniques

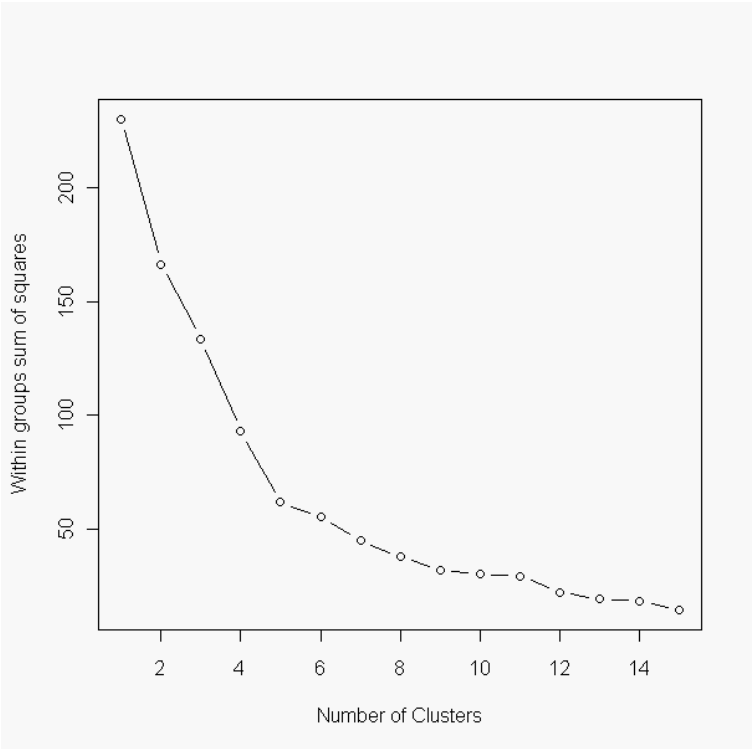
- Explore the scree plot



- A sharp bend occurs at end of the fourth segment.
- Start out with five clusters.

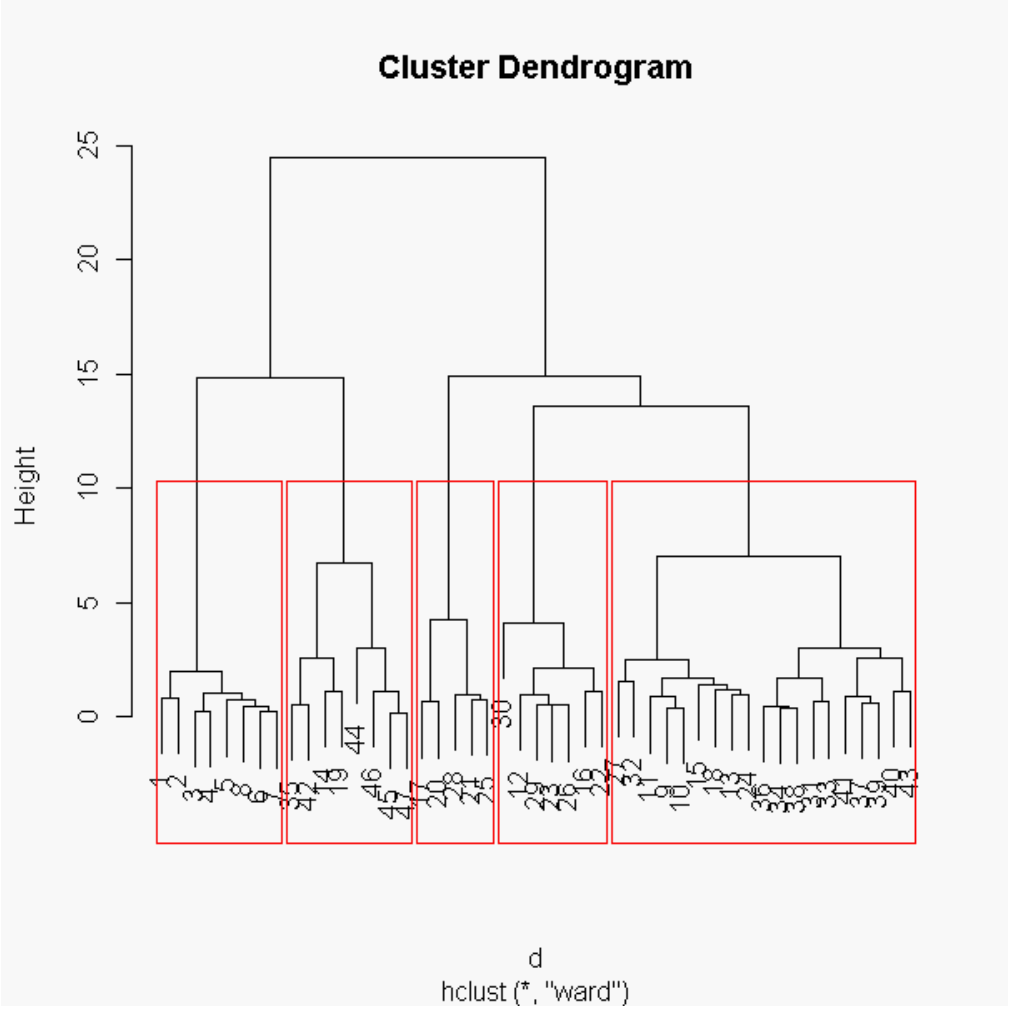
Clustering Techniques

- Explore the scree plot



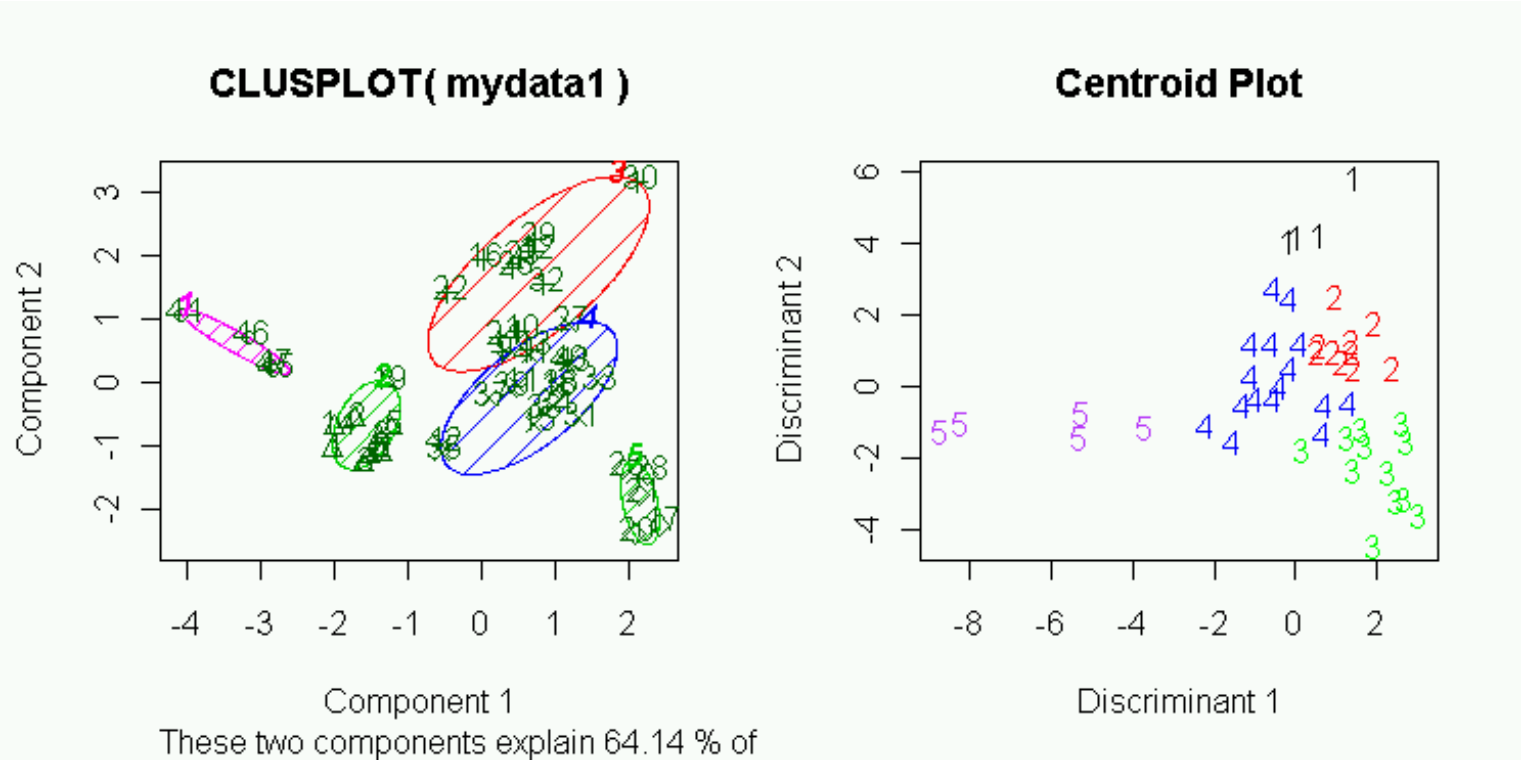
- A sharp bend occurs at end of the fourth segment.
- Start out with five clusters.

Ward Hierarchical & Cluster Dendogram



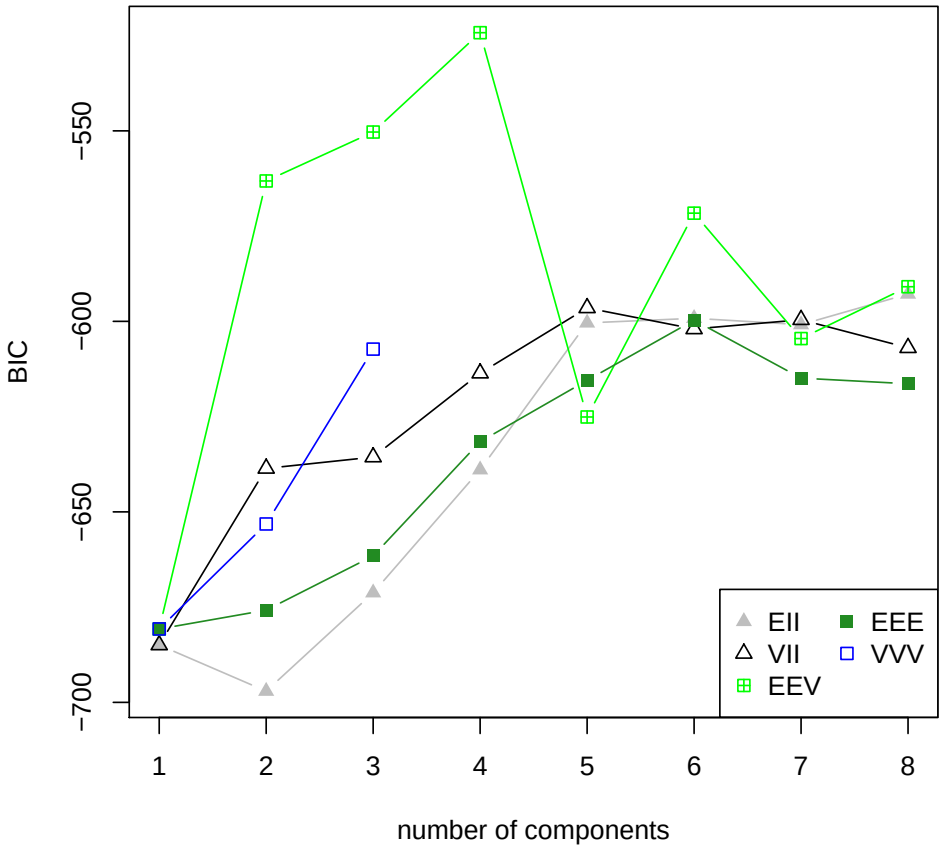
K-means clustering

- Five clusters are pretty well separated



- K-means don't work well in the presence of different size clusters.

Model-Based Clustering



Clustering validation

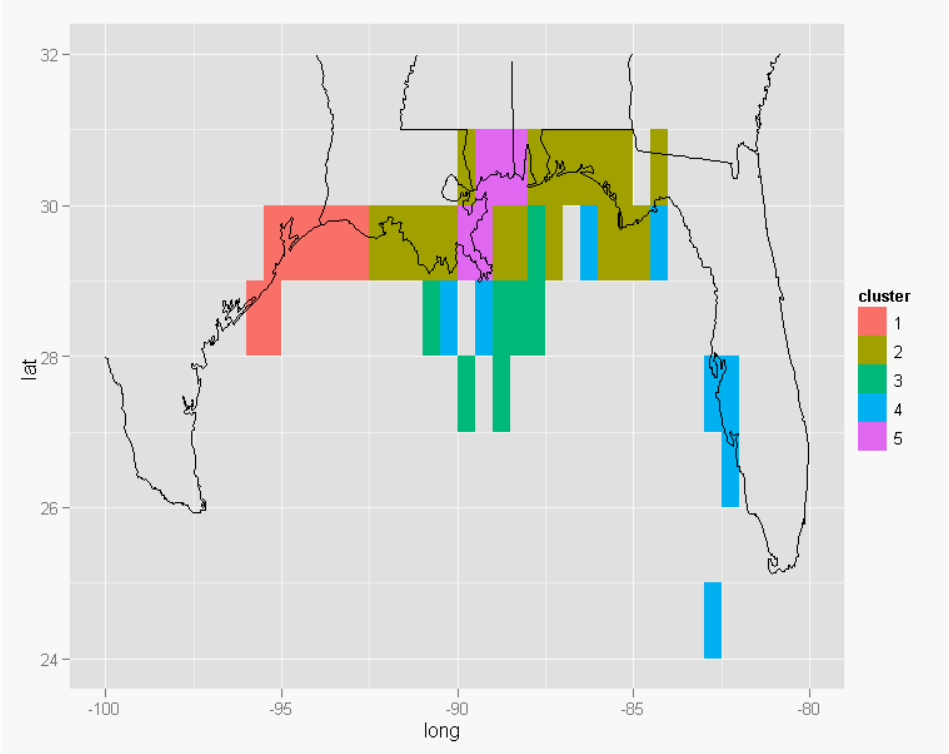
- Ward Method versus Model EEV

Criterion	Ward	model EEV	Better if
1	1121.62	1478.78	Lower
2	1.05	0.56	Higher
3	0.00	0.00	Lower

- Ward performs better than model EEV.

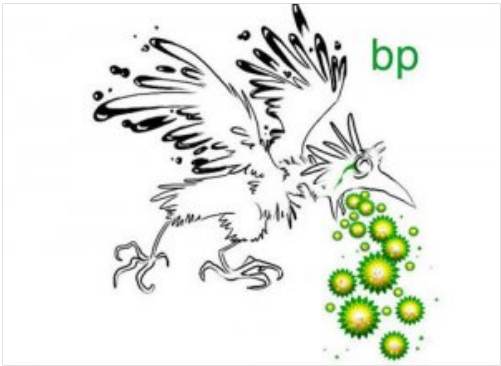
Spatial Mapping

- Map of the five clusters of based on combination of percent of oil and percent of death



Summary

- Death Percentage increases in data over 27 weeks
- Oiled Percentage reaches 50% and remains consistent until well is capped
- Geographic areas can be clustered by similar oiled and death percentages
- Data limitations

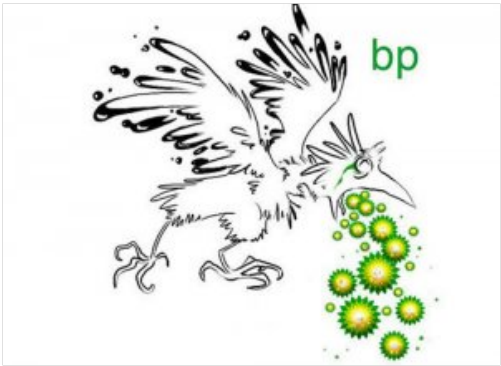


Thanks!

Questions?

Summary

- Death Percentage increases in data over 27 weeks
- Oiled Percentage reaches 50% and remains consistent until well is capped
- Geographic areas can be clustered by similar oiled and death percentages
- Data limitations

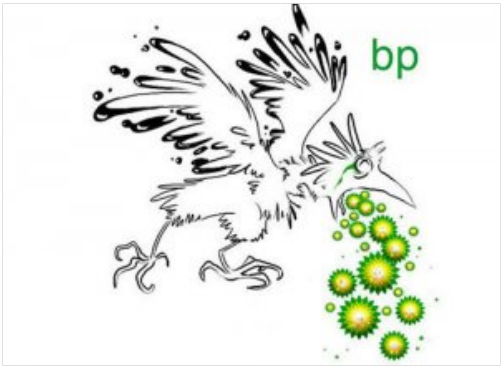


Thanks!

Questions?

Summary

- Death Percentage increases in data over 27 weeks
- Oiled Percentage reaches 50% and remains consistent until well is capped
- Geographic areas can be clustered by similar oiled and death percentages
- Data limitations

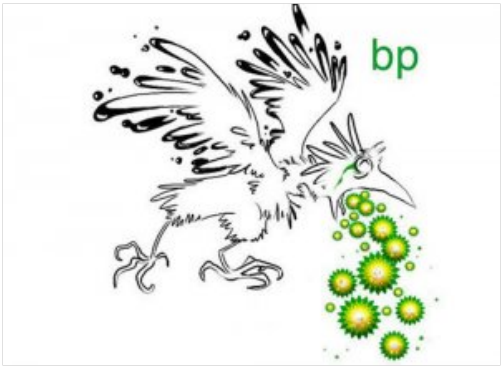


Thanks!

Questions?

Summary

- Death Percentage increases in data over 27 weeks
- Oiled Percentage reaches 50% and remains consistent until well is capped
- Geographic areas can be clustered by similar oiled and death percentages
- Data limitations

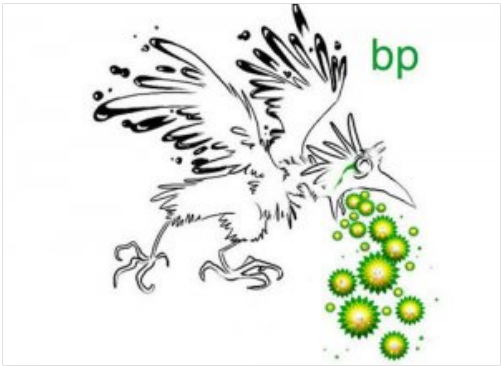


Thanks!

Questions?

Summary

- Death Percentage increases in data over 27 weeks
- Oiled Percentage reaches 50% and remains consistent until well is capped
- Geographic areas can be clustered by similar oiled and death percentages
- Data limitations



Thanks!

Questions?