



Society of Actuaries in Ireland

Graphing / Visualisation with R

27-11-2017

Disclaimer

**The views expressed in these presentations
are those of the presenter(s) and not
necessarily of the Society of Actuaries in
Ireland**

Welcome

- Pedro Ecija Serrano
Chair, Data Analytics Subcommittee



Disclaimer:

The material, content and views in the following presentation are those of the presenter(s).

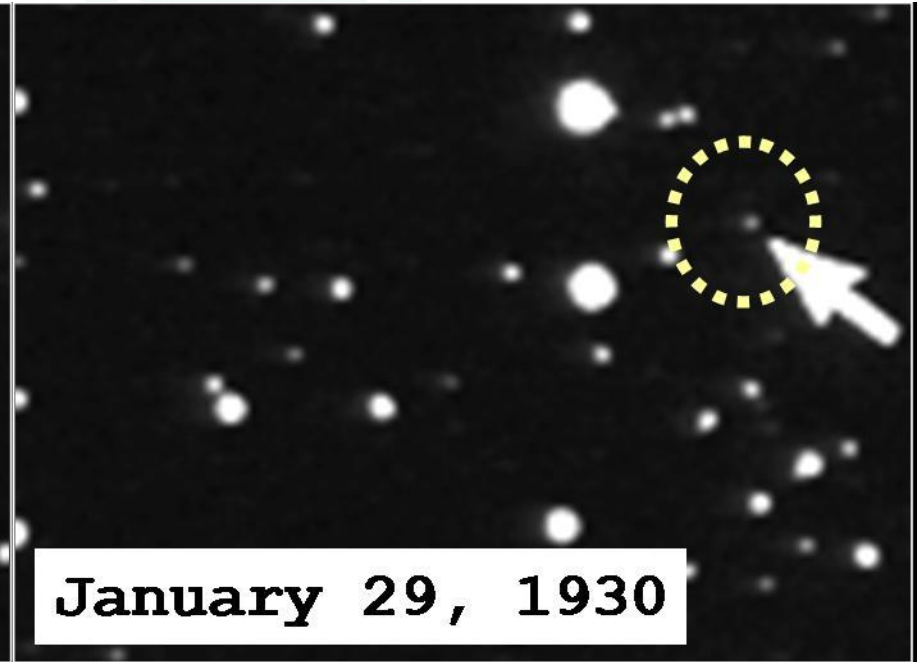
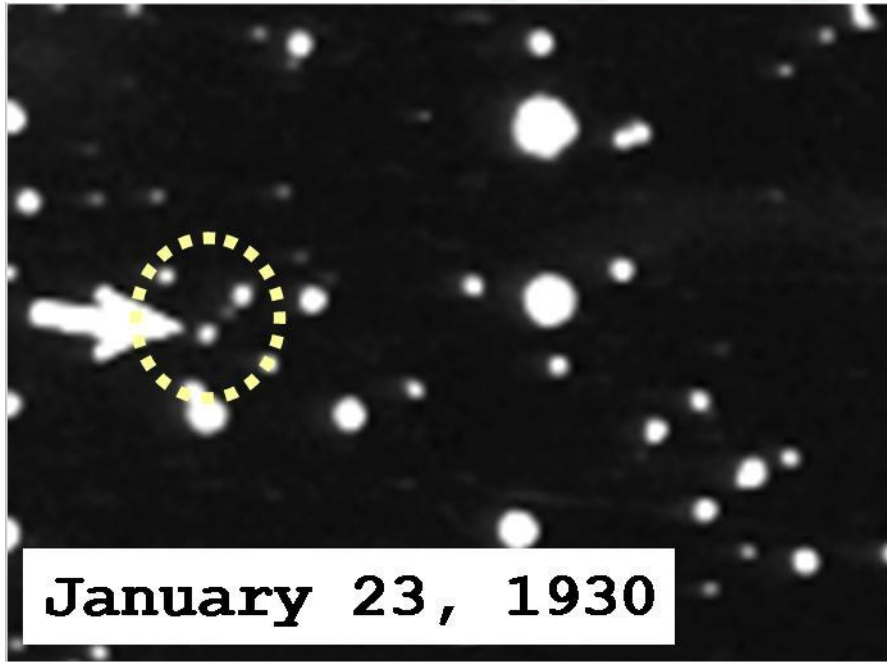
Agenda

- Introduction
- The case for graphical investigation
- Is Excel the best tool?
- Possibilities with (base) R
- Basic demonstrations
- Advanced demonstrations
- Wrap up
- Questions

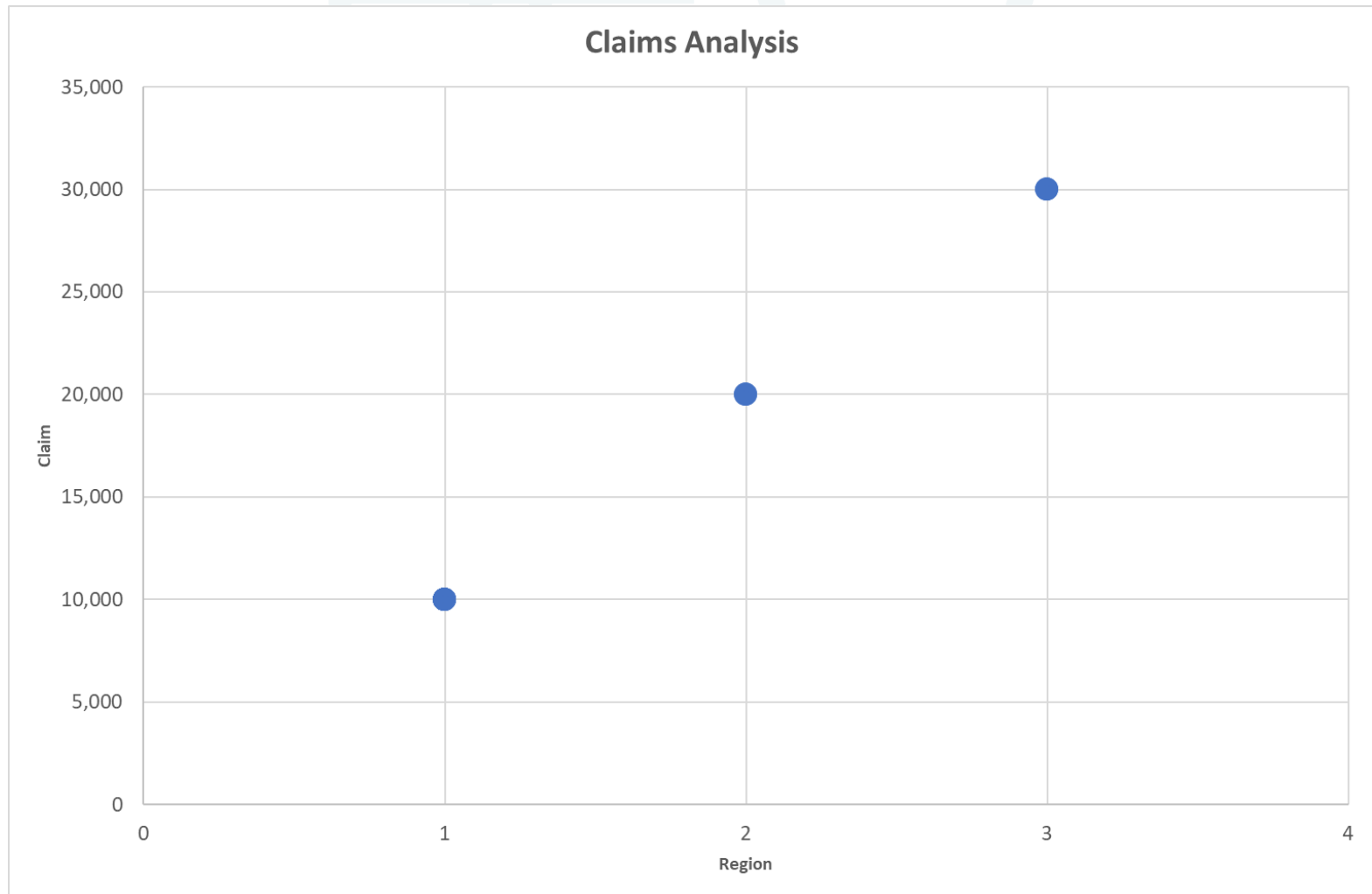
Introduction

- Desktop tools
- Not ML systems
- Not all presentation friendly
- Some demonstrations
- Datasets illustrative
- Not considering analytical overlays
- Code will be made available
- Beginner to Intermediate level

The case for graphical investigation



Example 1 - Excel



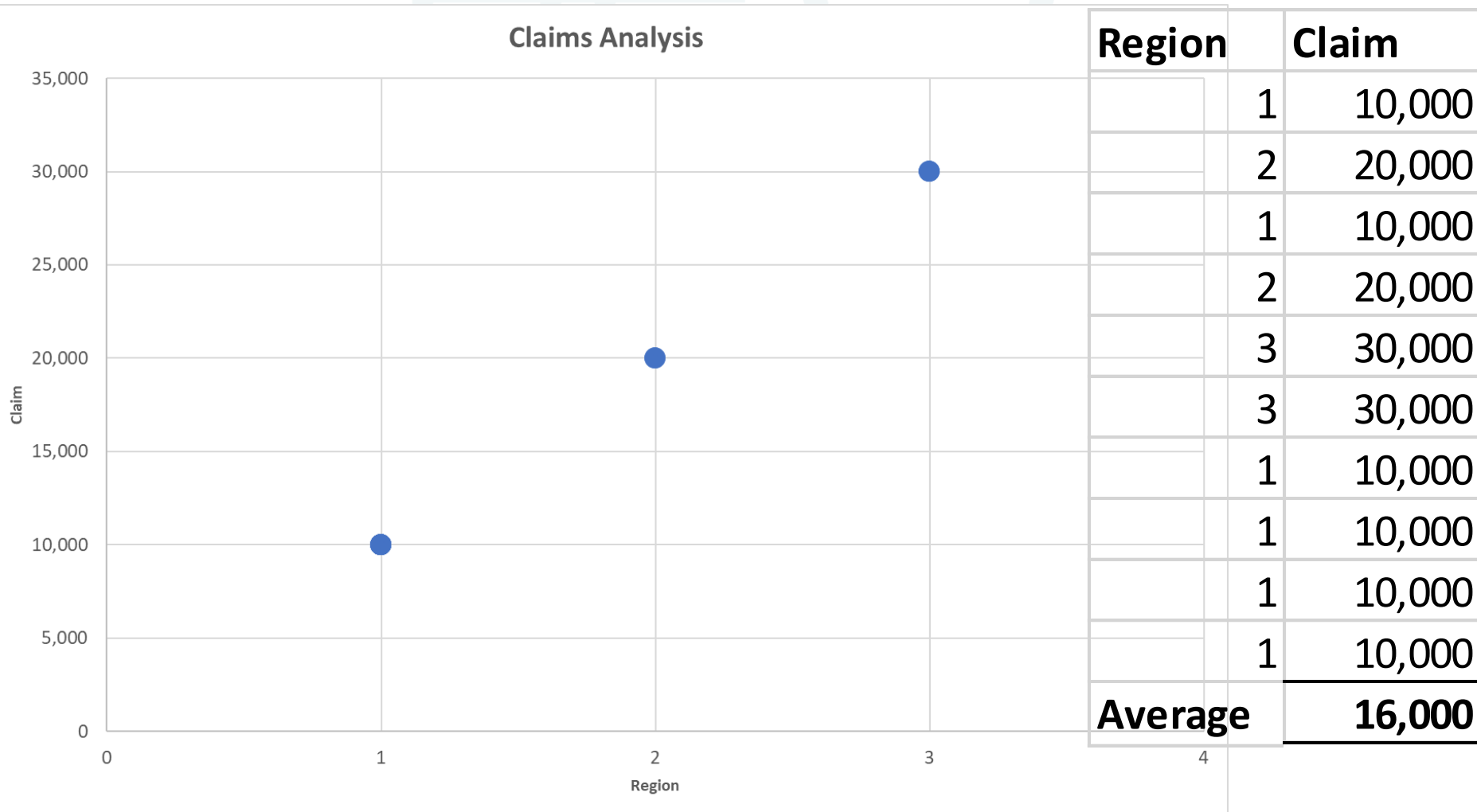
What's the average claim?

Claim dataset

Region	Claim
1	10,000
2	20,000
1	10,000
2	20,000
3	30,000
3	30,000
1	10,000
1	10,000
1	10,000
1	10,000
Average	16,000

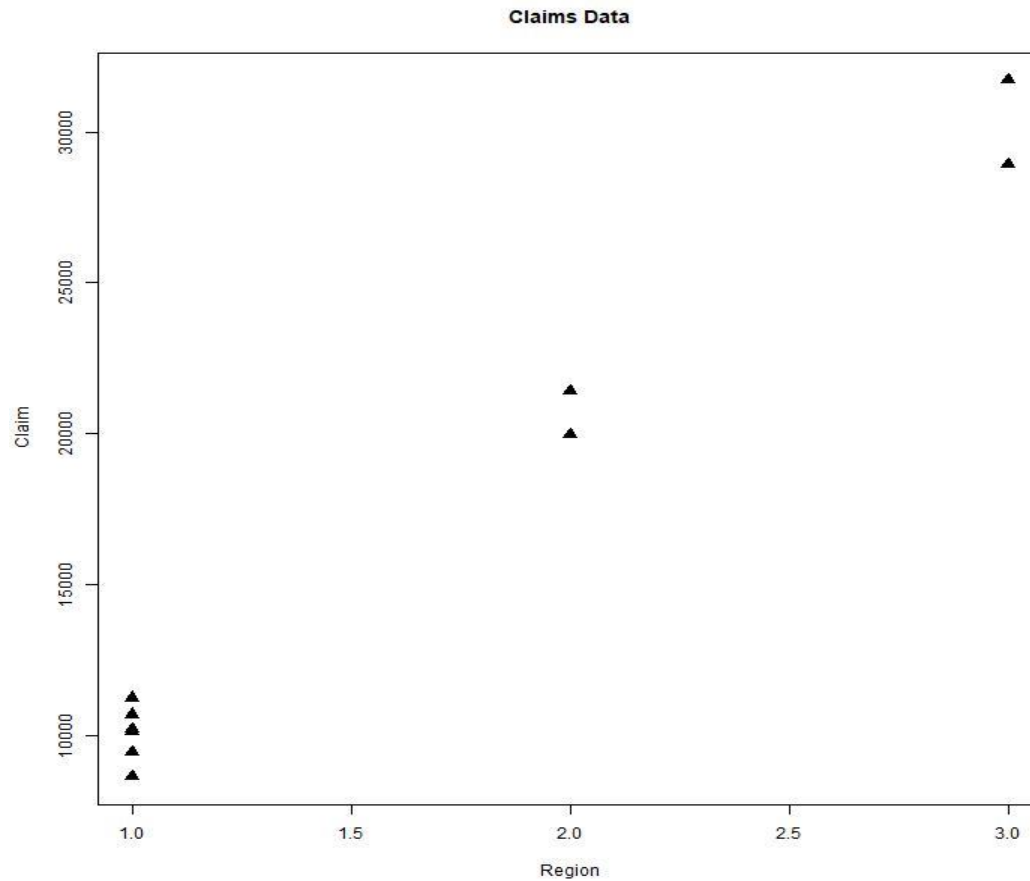
What's the average claim now?

Example 1 - Excel



What's the average claim now?

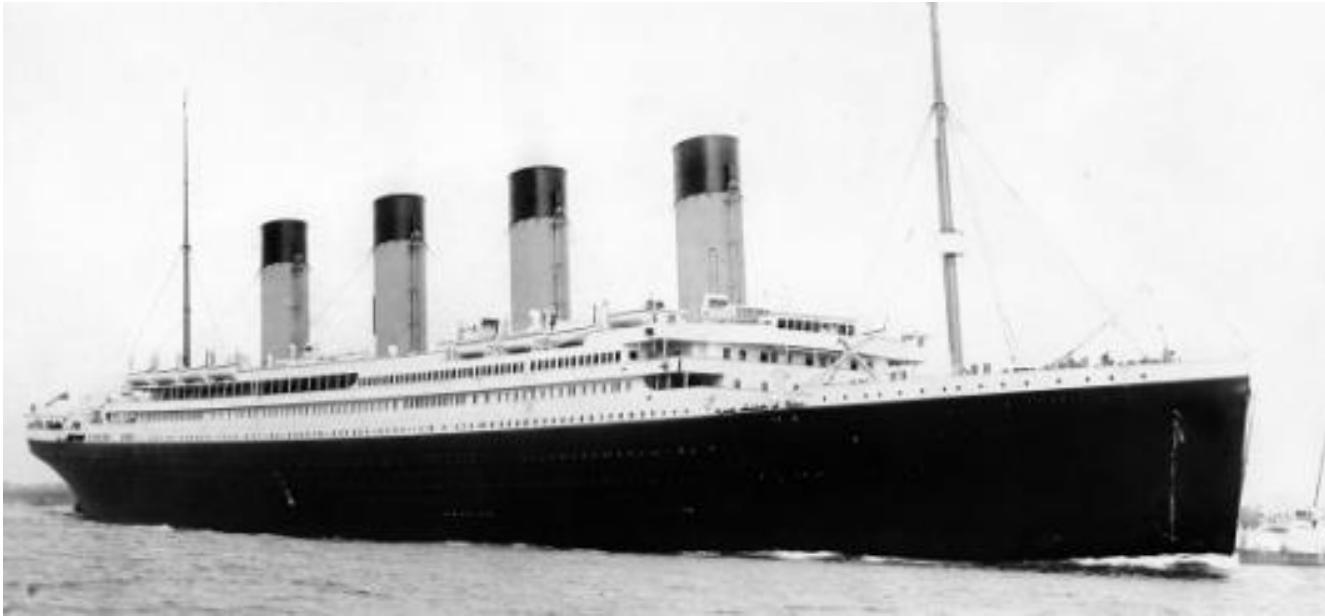
Example 1 - R



What's the average claim now?



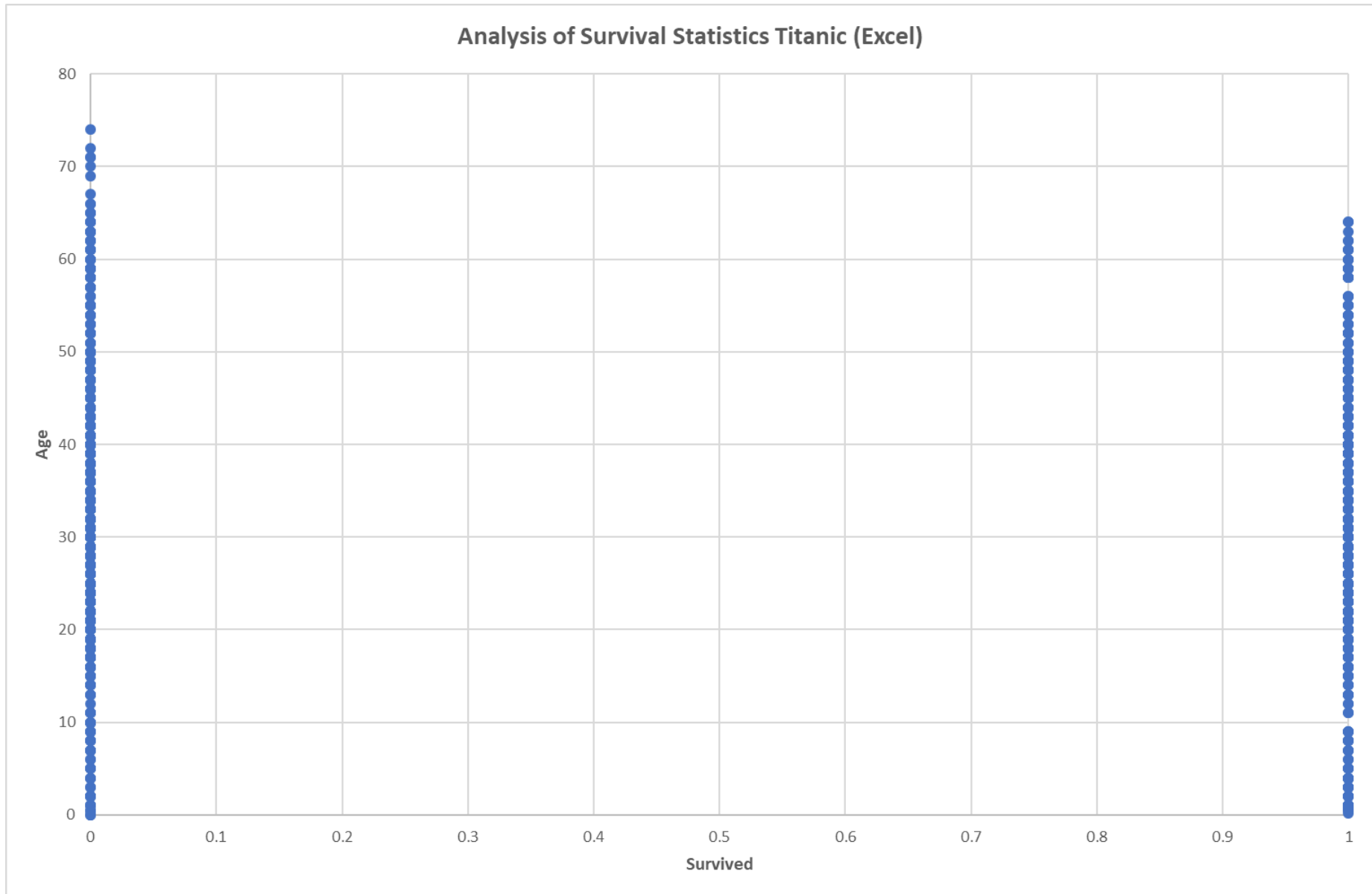
Titanic Dataset



	name	gender	age	class	embarked	country	ticketno	fare	sibsp	parch	survived
1	Abbing, Mr. Anthony	male	42	3	S	United States	5547	7	0	0	0
2	Abbott, Mr. Eugene Joseph	male	13	3	S	United States	2673	20	0	2	0
3	Abbott, Mr. Rossmore Edward	male	16	3	S	United States	2673	20	1	1	0
4	Abbott, Mrs. Rhoda Mary 'Rosa'	female	39	3	S	England	2673	20	1	1	1
5	Abelseth, Miss. Karen Marie	female	16	3	S	Norway	348125	7	0	0	1
:	:	:	:	:	:	:	:	:	:	:	:
2203	Wynn, Mr. Walter	male	41	4	B	England	NA	NA	NA	NA	1
2204	Yearsley, Mr. Harry	male	40	4	S	England	NA	NA	NA	NA	1
2205	Young, Mr. Francis James	male	32	4	S	England	NA	NA	NA	NA	0
2206	Zanetti, Sig. Minio	male	20	4	S	England	NA	NA	NA	NA	0
2207	Zarracchi, Sig. L.	male	26	4	S	England	NA	NA	NA	NA	0

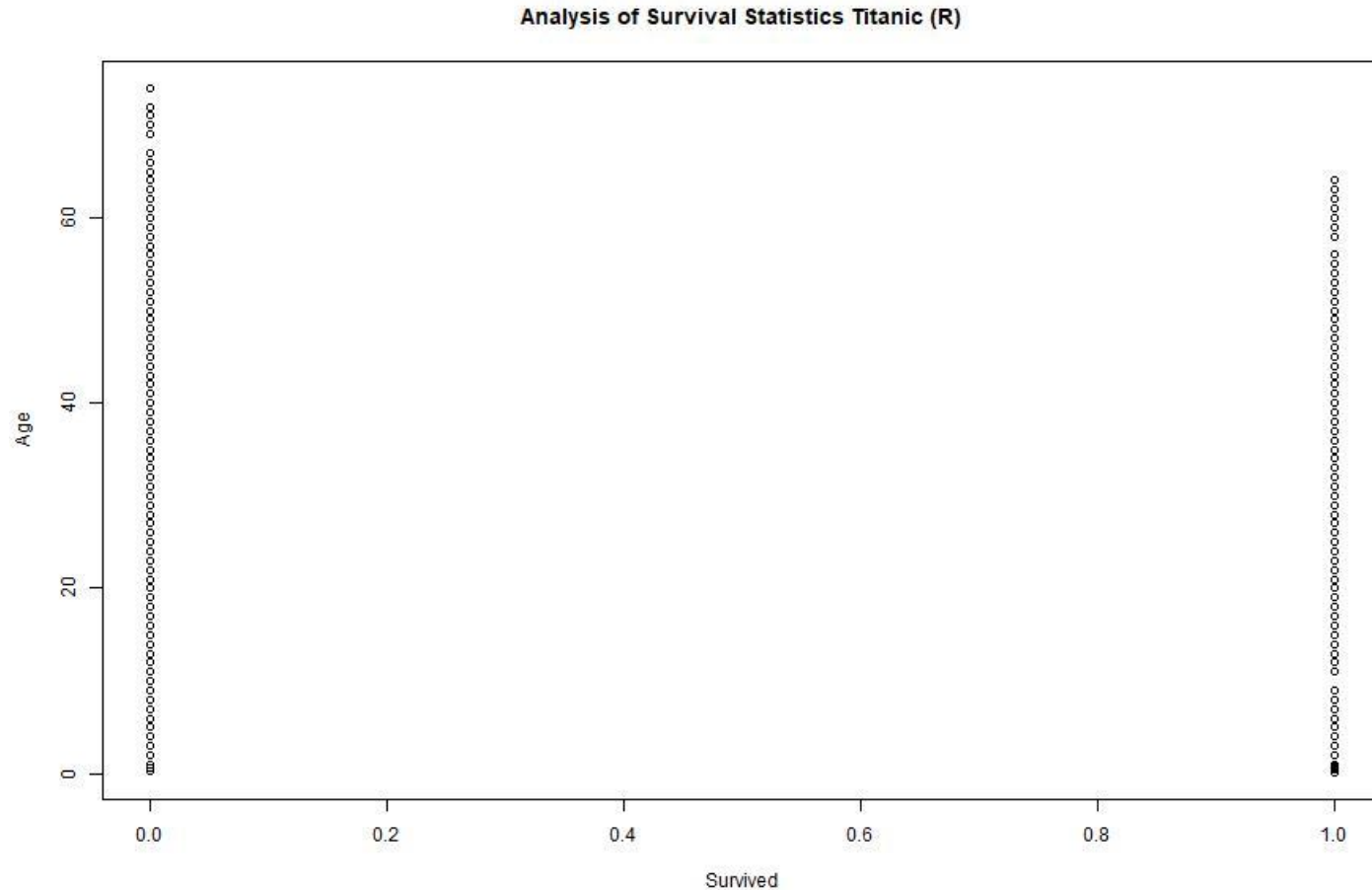


Titanic Data – Poor presentation (EXCEL)





Titanic Data – Poor presentation

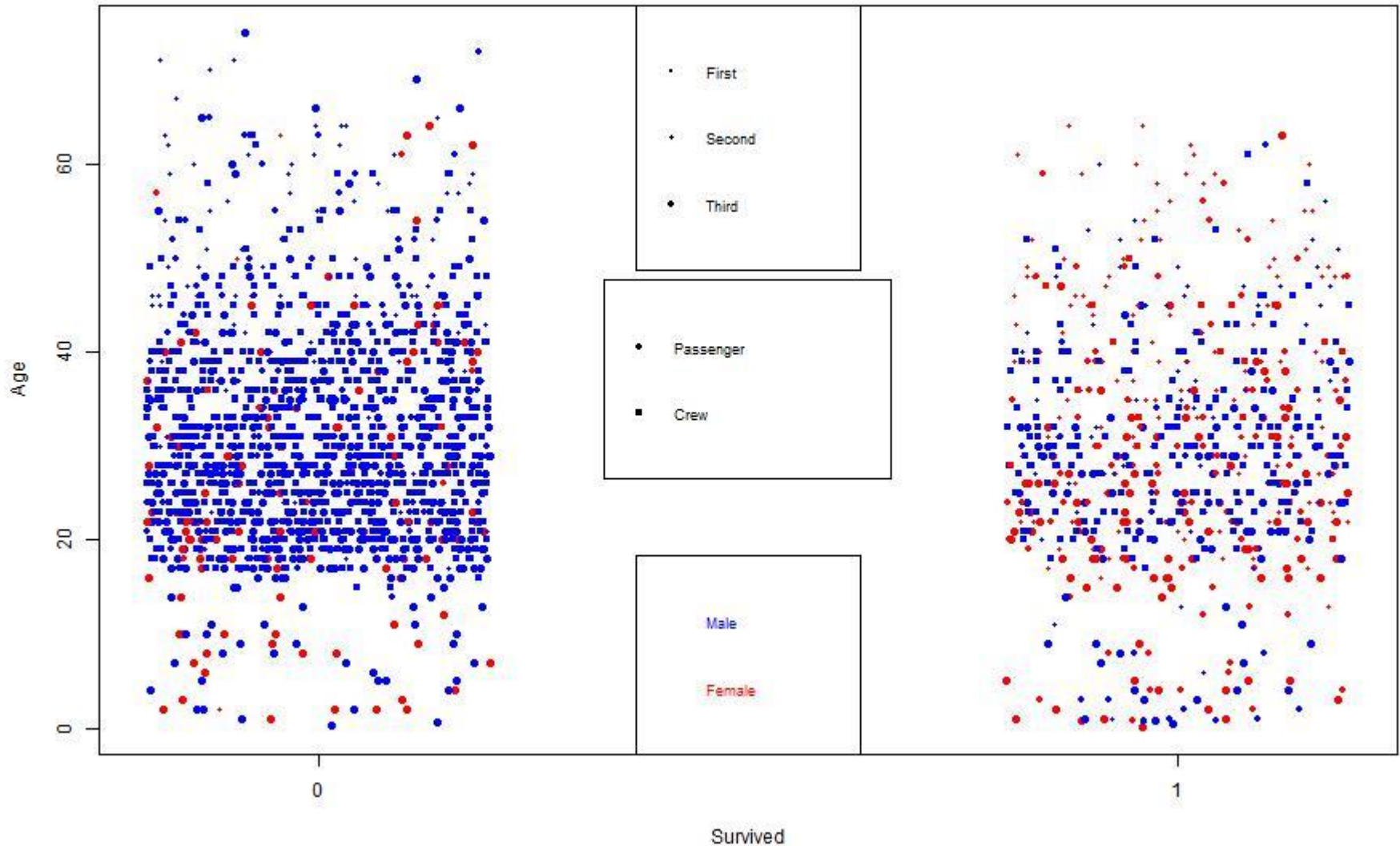


- Excellent analytical and graphing tools
- R base graph (~EXCEL)



Titanic Data – Superior presentation

Analysis of Survival Statistics Titanic (R)





Rotatable & Zoomable charts

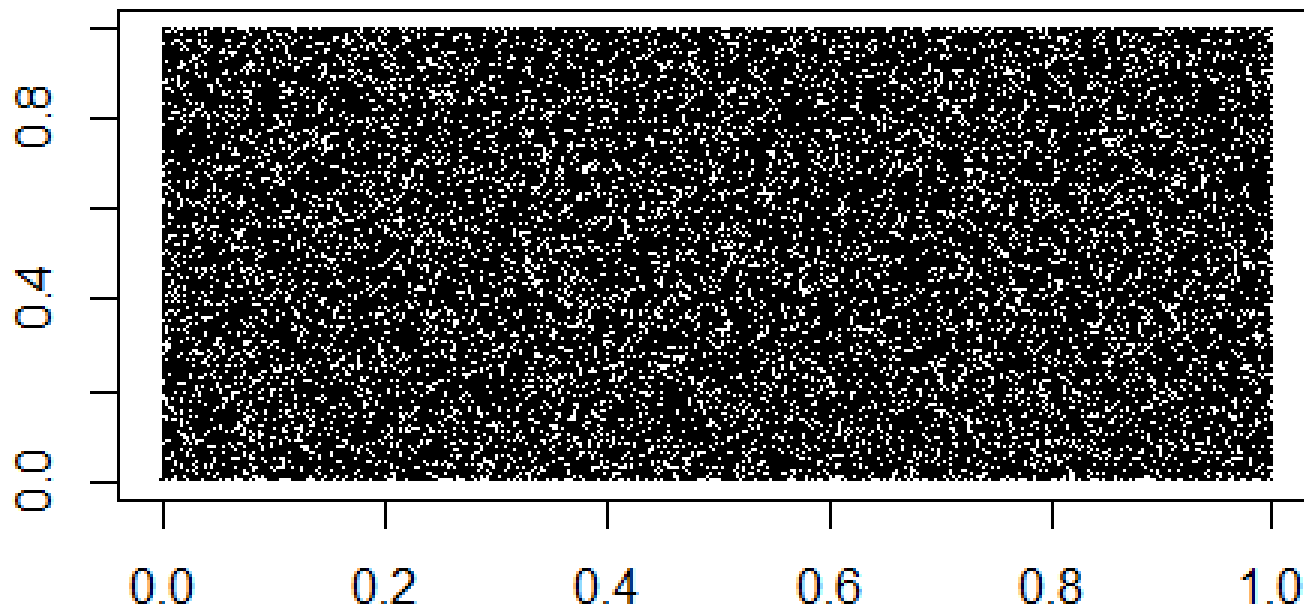
Demonstrate using Titanic dataset



Random number generation 1

(In)famous RANDU sequence

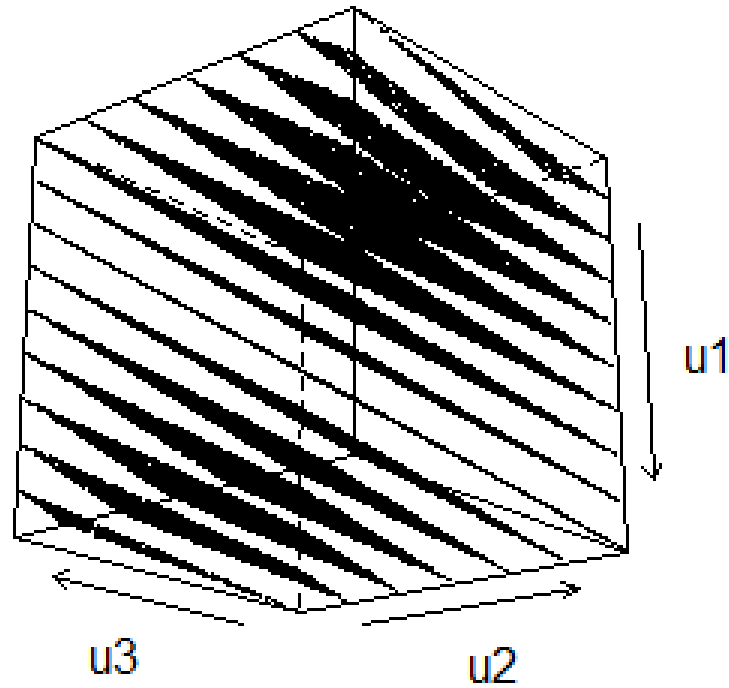
$$X_{t+1} = 65539 * X_t \bmod 2^{31}$$



IBM's RANDU is widely considered to be one of the most ill-conceived random number generators ever designed



Random number generation 2





Animated graphics

	country	continent	year	lifeExp	population	gdpPercap
1	Afghanistan	Asia	1952	28.8	8,425,333	779
2	Afghanistan	Asia	1957	30.3	9,240,934	821
3	Afghanistan	Asia	1962	32.0	10,267,083	853
4	Afghanistan	Asia	1967	34.0	11,537,966	836
5	Afghanistan	Asia	1972	36.1	13,079,460	740
6	Afghanistan	Asia	1977	38.4	14,880,372	786
7	Afghanistan	Asia	1982	39.9	12,881,816	978
8	Afghanistan	Asia	1987	40.8	13,867,957	852
9	Afghanistan	Asia	1992	41.7	16,317,921	649
10	Afghanistan	Asia	1997	41.8	22,227,415	635
11	Afghanistan	Asia	2002	42.1	25,268,405	727
12	Afghanistan	Asia	2007	43.8	31,889,923	975
13	Albania	Europe	1952	55.2	1,282,697	1,601
14	Albania	Europe	1957	59.3	1,476,505	1,942
15	Albania	Europe	1962	64.8	1,728,137	2,313
16	Albania	Europe	1967	66.2	1,984,060	2,760

See <https://www.youtube.com/watch?v=jbkSRLYSojo>

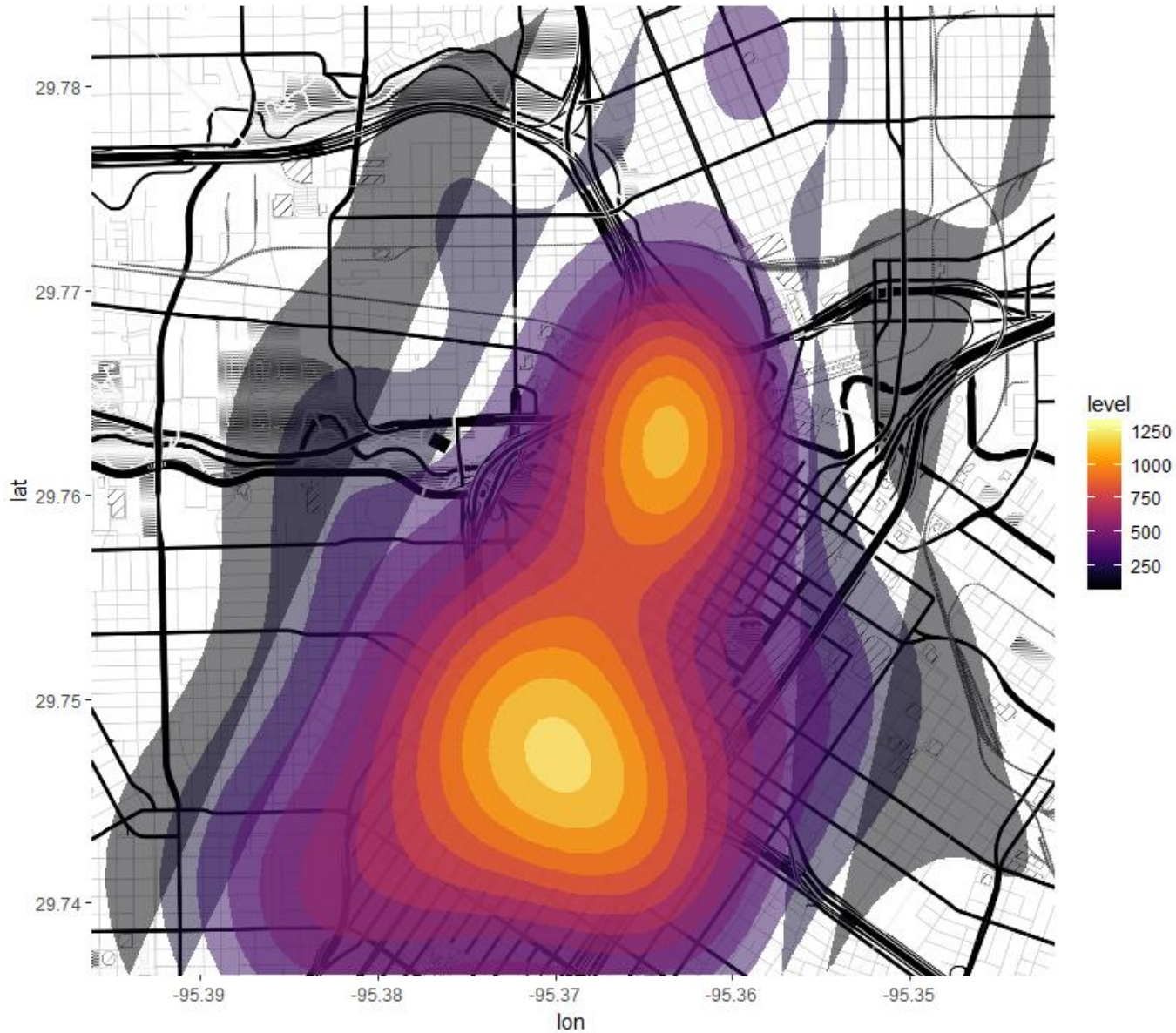


Crime statistics and mapping 1



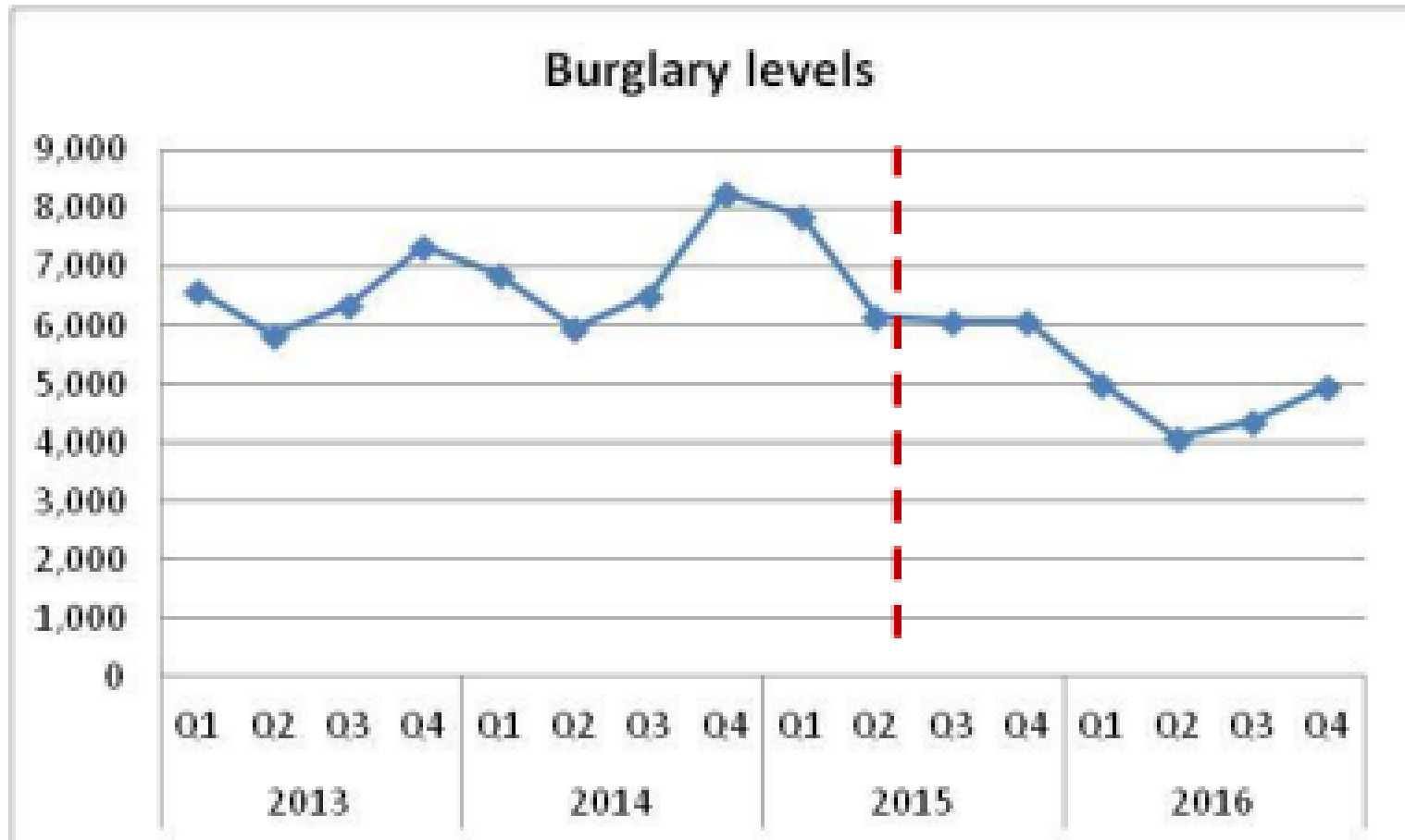


Crime statistics and mapping 2



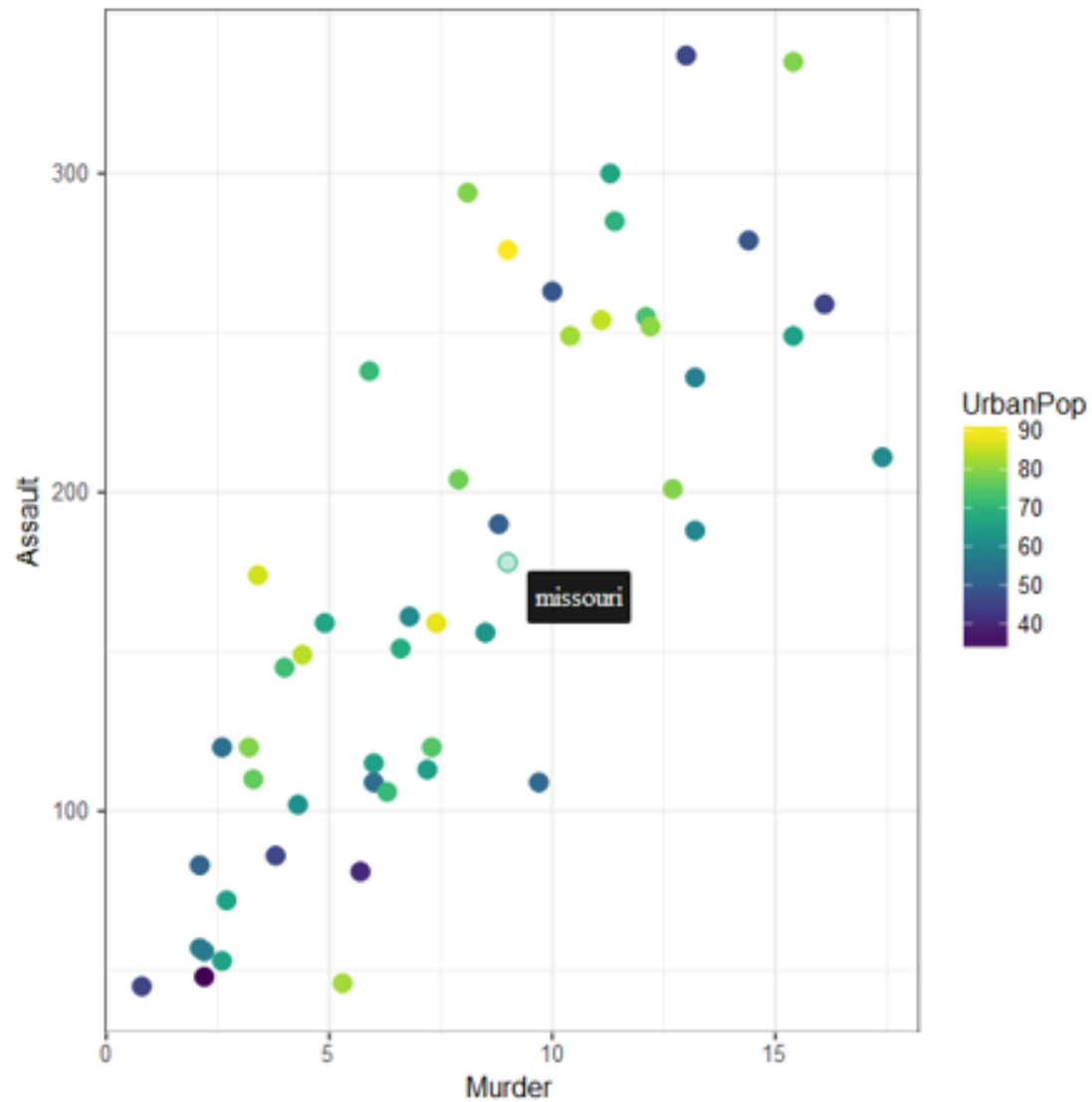


Crime statistics in Ireland



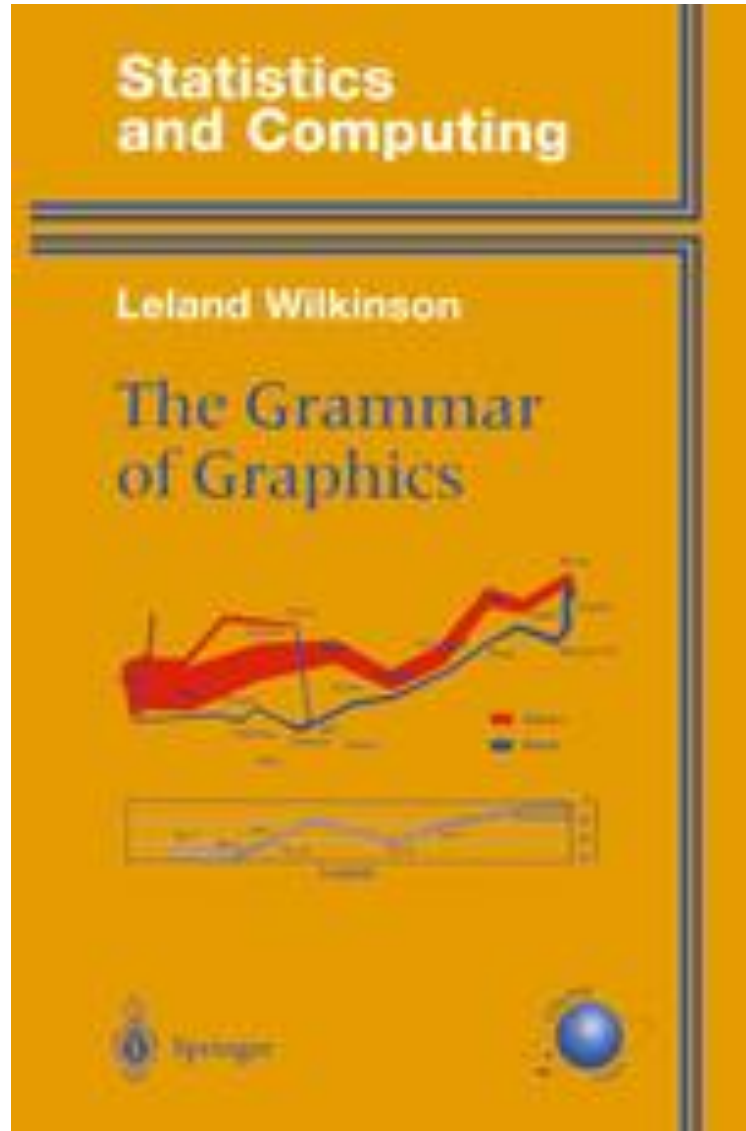


Clickable graphics





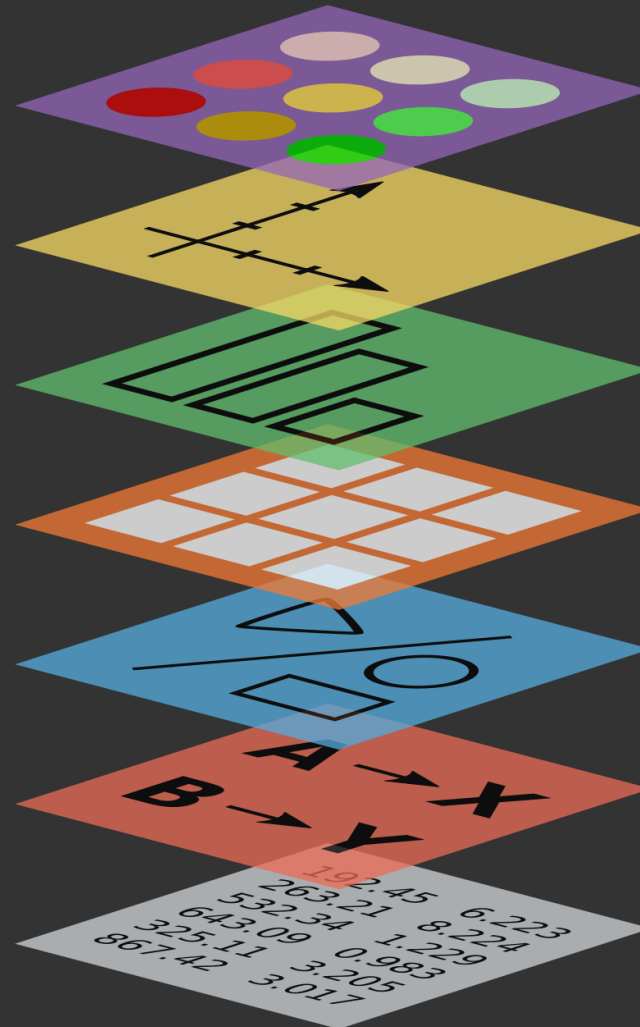
The grammar of graphics





The grammar of graphics 2 – ggplot2

Theme
Coordinates
Statistics
Facets
Geometries
Aesthetics
Data





Detailed demonstrations



Chorddiag & Circlize

- Taking another look at the Titanic Dataset

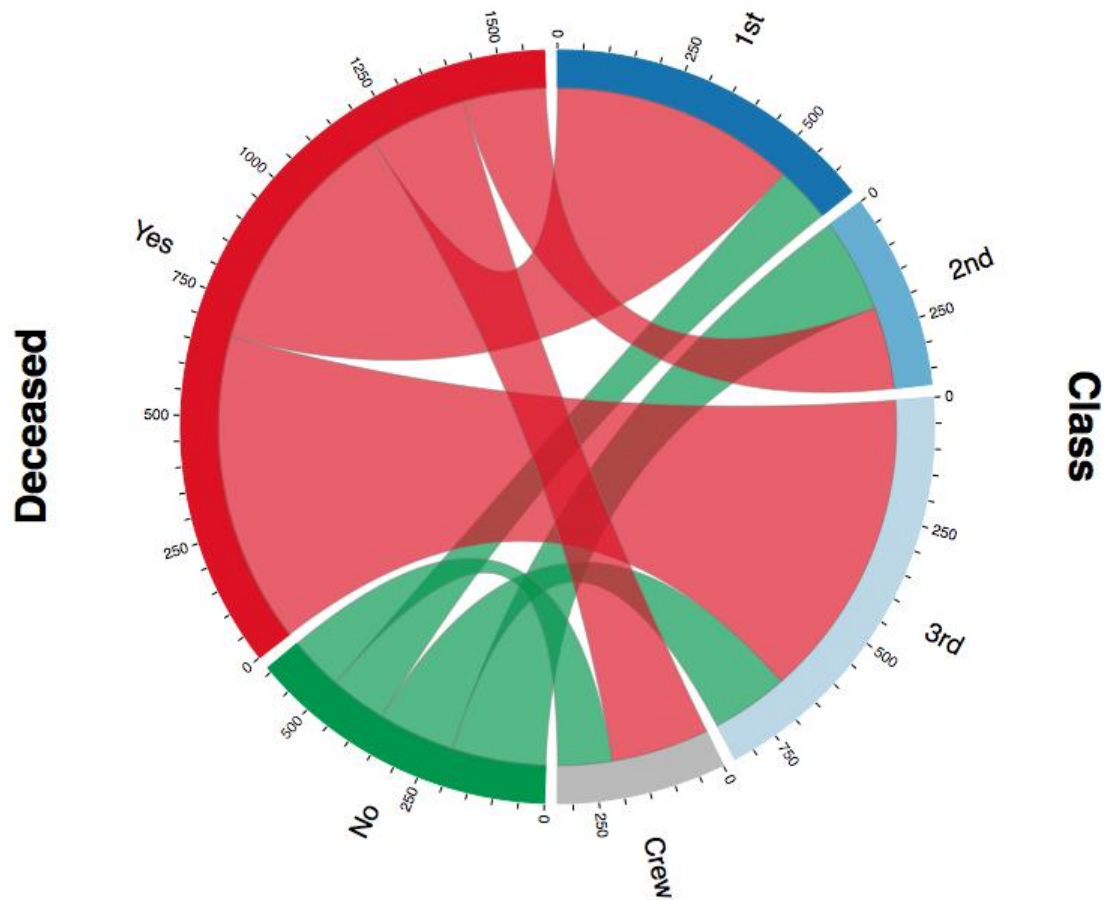
Ordinal
Data

Class	Sex	Age	Survived	n
1st	Male	Child	No	0
2nd	Male	Child	No	0
3rd	Male	Child	No	35
Crew	Male	Child	No	0
1st	Female	Child	No	0
2nd	Female	Child	No	0
3rd	Female	Child	No	17
Crew	Female	Child	No	0
1st	Male	Adult	No	118
2nd	Male	Adult	No	154
3rd	Male	Adult	No	387
Crew	Male	Adult	No	670
1st	Female	Adult	No	4
2nd	Female	Adult	No	13
3rd	Female	Adult	No	89
Crew	Female	Adult	No	3
1st	Male	Child	Yes	5
2nd	Male	Child	Yes	11
3rd	Male	Child	Yes	13
Crew	Male	Child	Yes	0
1st	Female	Child	Yes	1
2nd	Female	Child	Yes	13
3rd	Female	Child	Yes	14
Crew	Female	Child	Yes	0
1st	Male	Adult	Yes	57
2nd	Male	Adult	Yes	14
3rd	Male	Adult	Yes	75
Crew	Male	Adult	Yes	192
1st	Female	Adult	Yes	140
2nd	Female	Adult	Yes	80
3rd	Female	Adult	Yes	76
Crew	Female	Adult	Yes	20

Binary



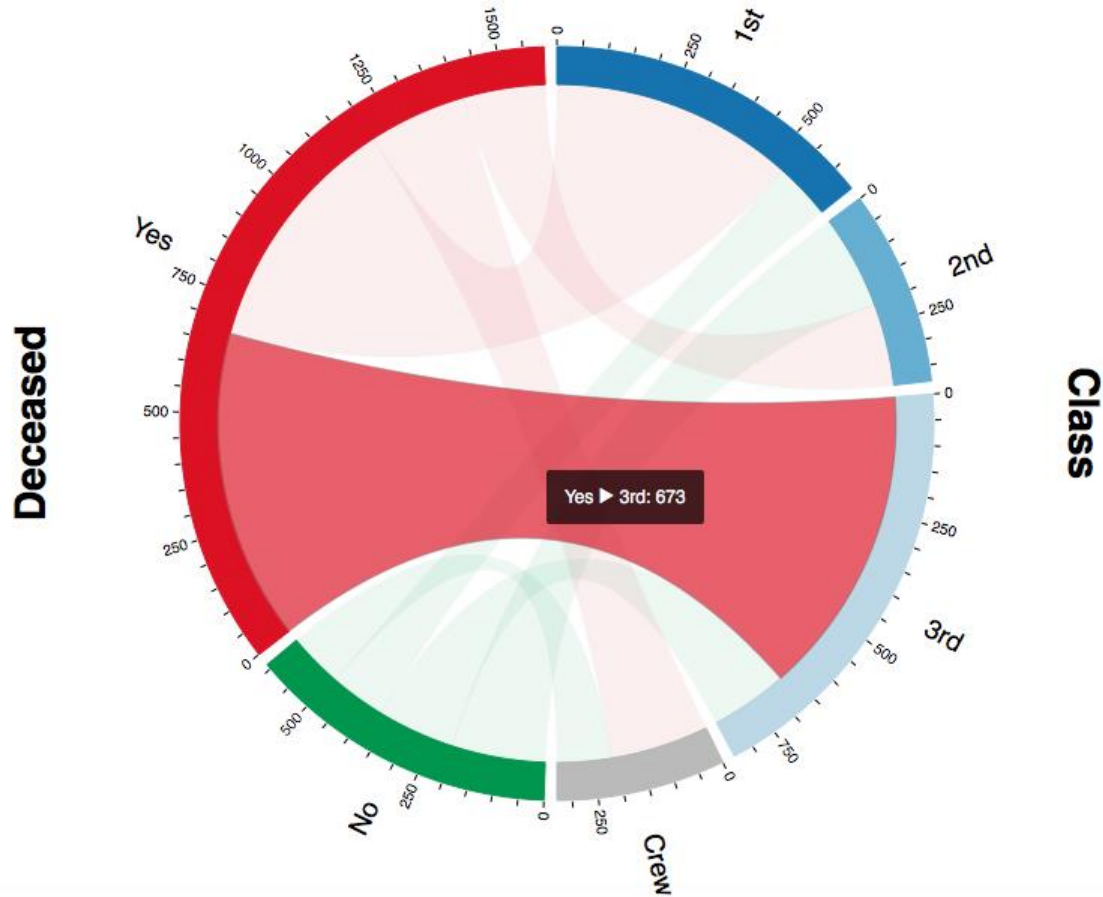
Chorddiag



```
chorddiag(titanic.mat, type = "bipartite", groupColors = groupColors, tickInterval = 50)
```

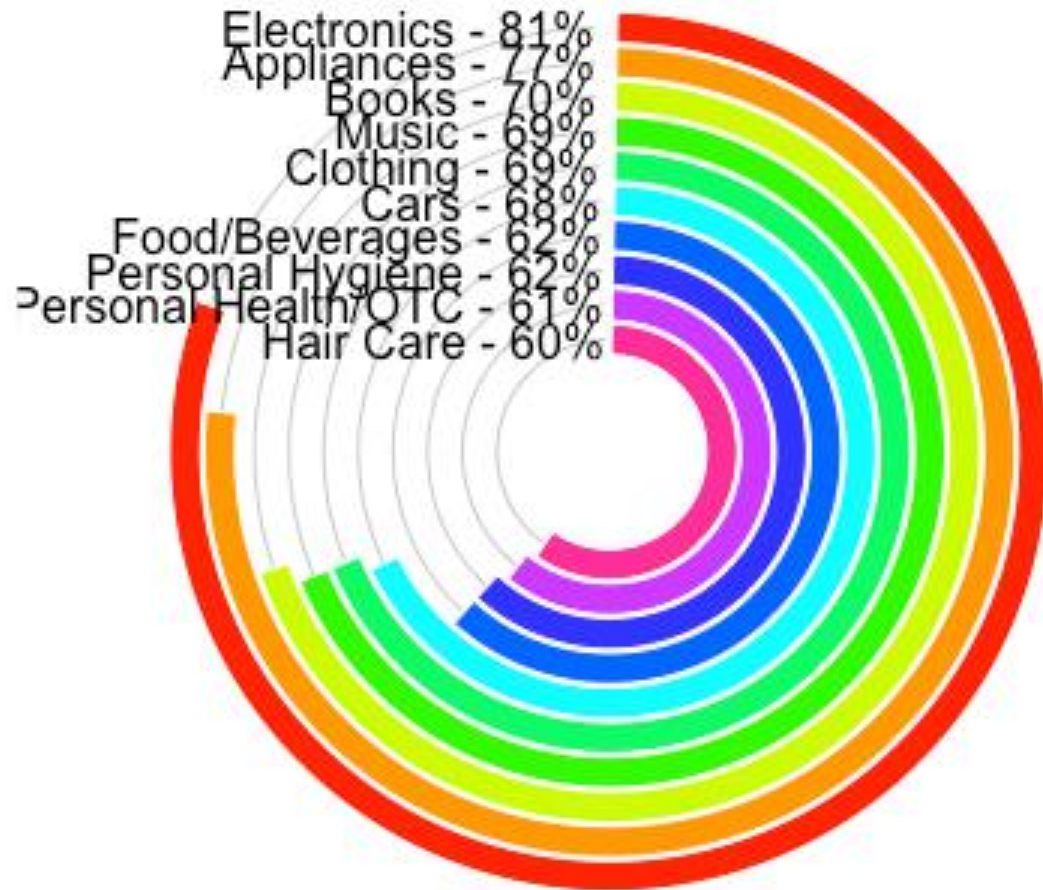


Chorddiag Interaction



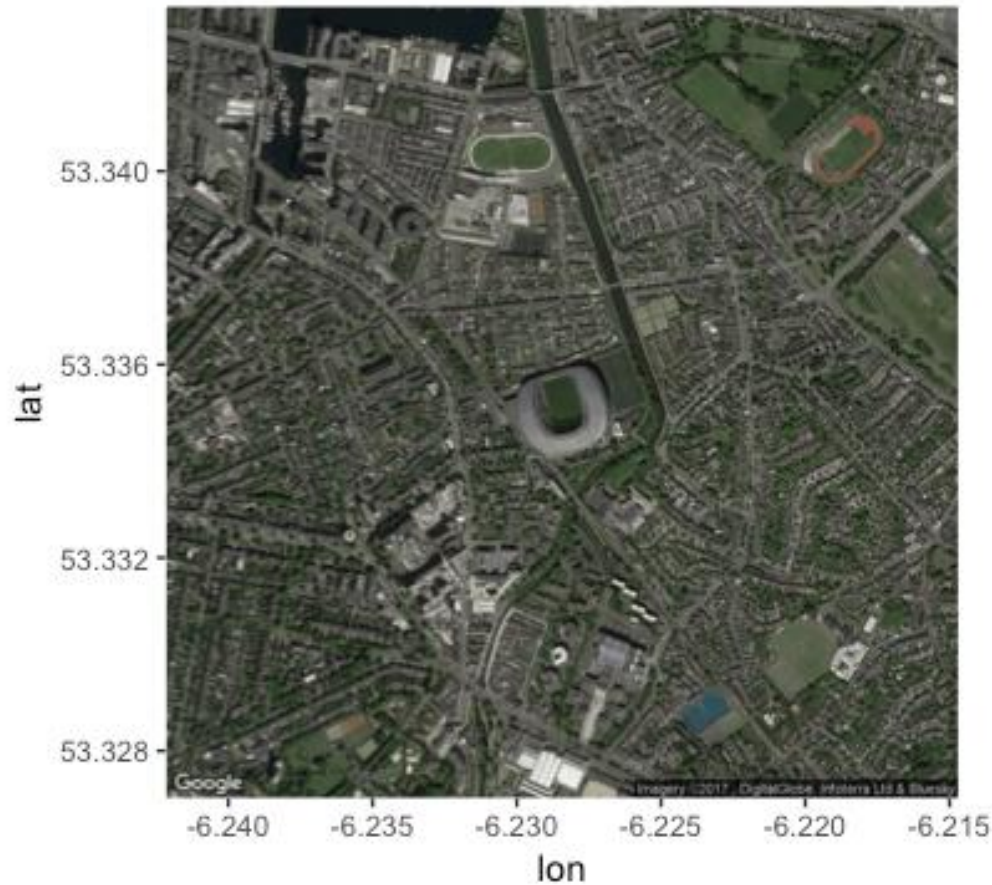


Circlize





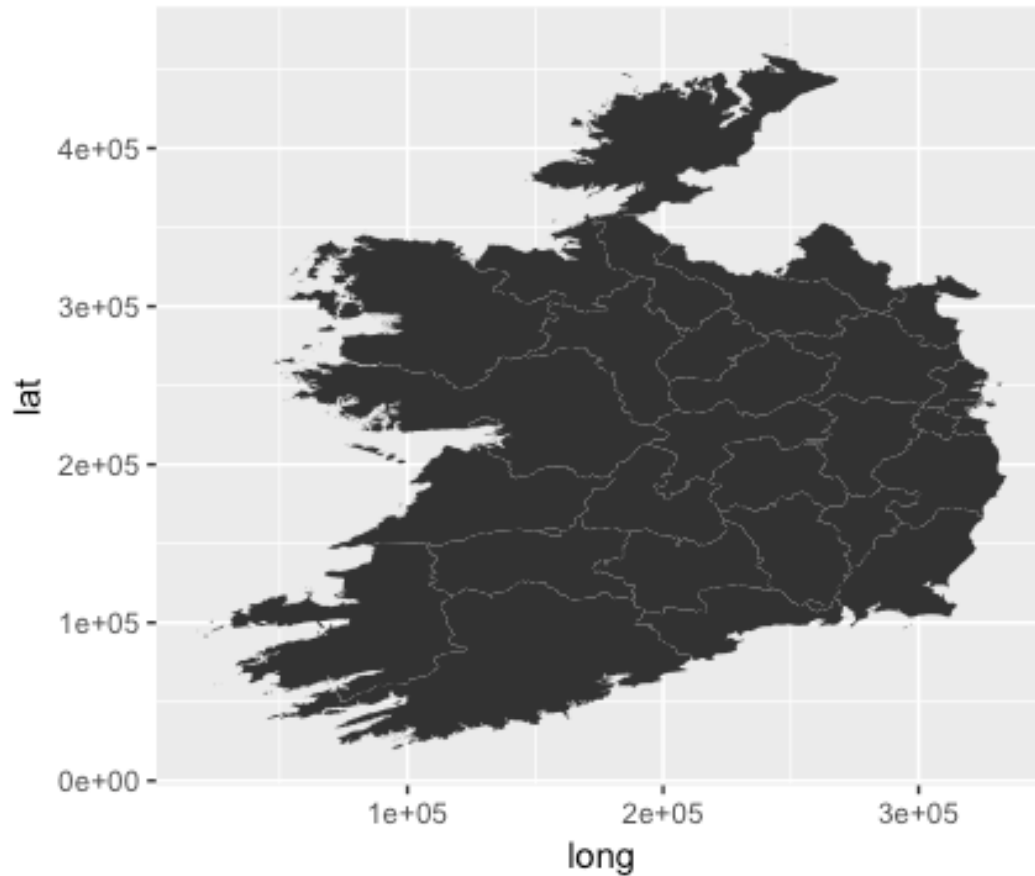
GGMAP



Accessing 'satellite' and 'hybrid' maps straight from Google Map's API using qmap



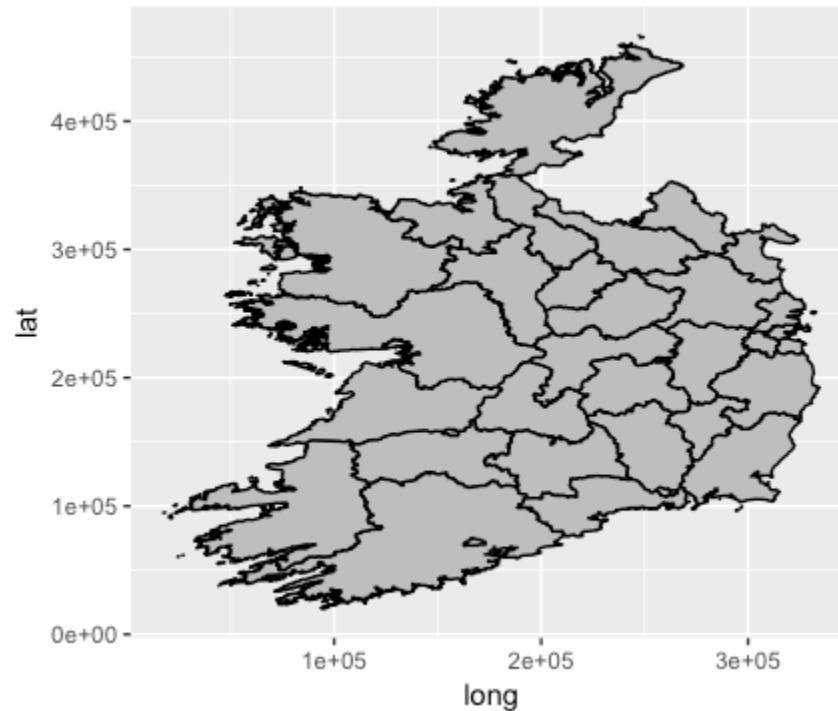
GGPLOT



```
basicmap <- ggplot(counties.df) + aes(long, lat, group=group) +  
geom_polygon()
```



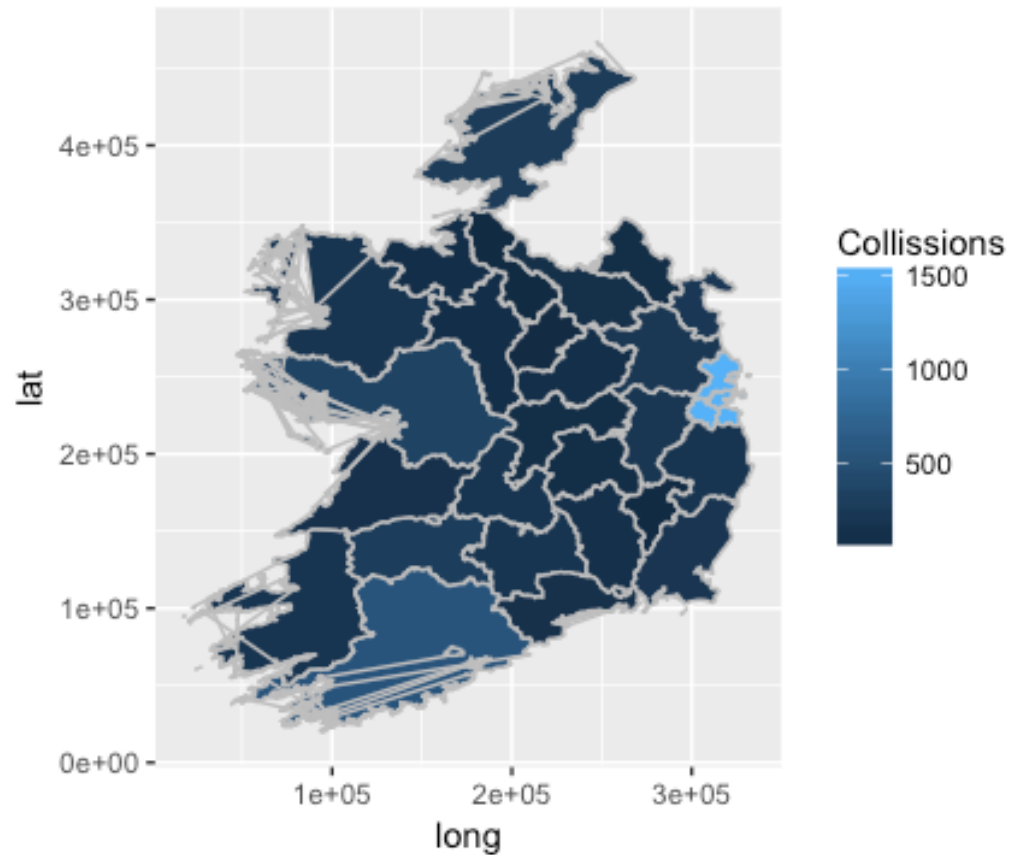
GGPLOT



```
countymap <- ggplot(counties.df) + aes(long, lat, group=group) +  
geom_polygon(colour="black", fill="grey")
```



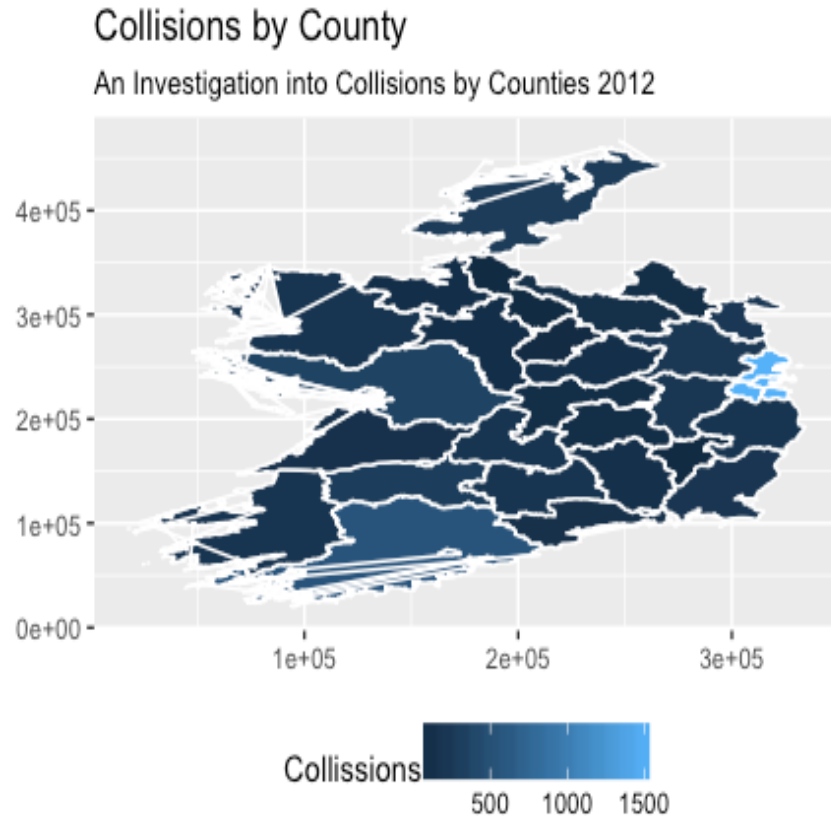

GGPLOT



Using inner join from Dplyr on two dataframes to overlap collision data(from CSO) over geospatial data



Iterations of GGPLOT



Author: Sean Cahill

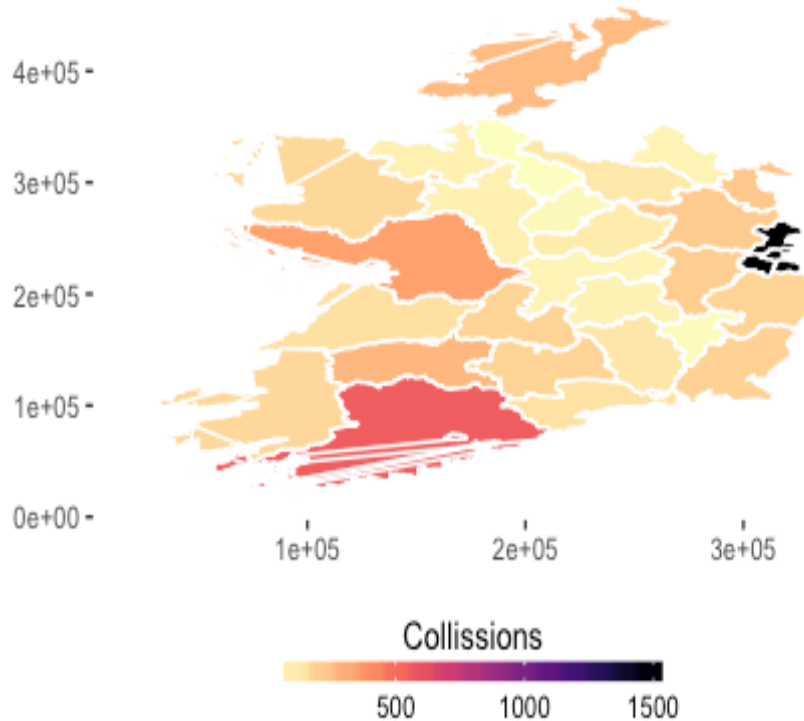
```
q<-p+theme(legend.position = "bottom")+....
```



Iterations of GGPLOT

Collisions by County

An Investigation into Collisions by Counties 2012



Author: Sean Cahill

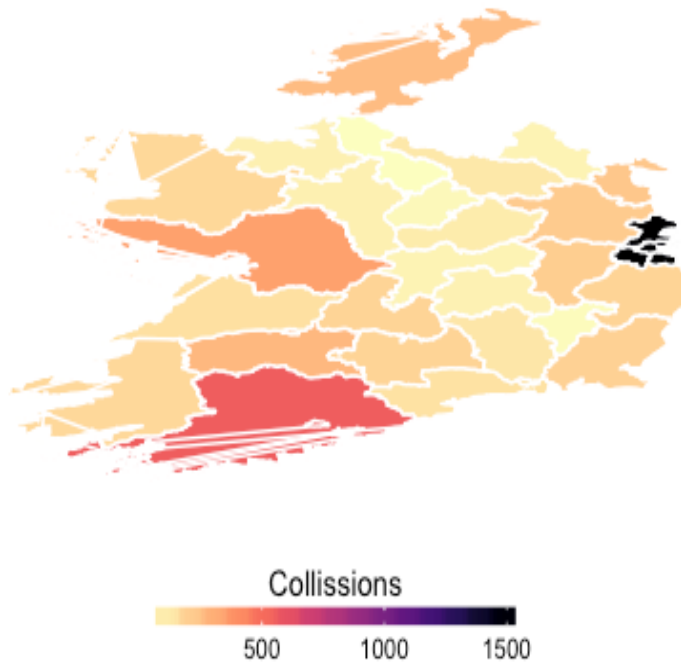
```
r<-q+scale_fill_viridis(option = "magma", direction = -1)....
```



Iterations of GGPLOT

Collisions by County

An Investigation into Collisions by Counties 2012



Author: Sean Cahill

```
t<-r+theme(axis.title.x=element_blank(), axis.text.x=element_blank(),
```



3d Graphics-Plotly and Shiny

Simple steps to get started with Plotly

1. `install.packages("plotly")`
2. `Require(plotly)`
3. `Sys.setenv("plotly_username"="your_plotly_username")`

`Sys.setenv("plotly_api_key"="your_api_key")`

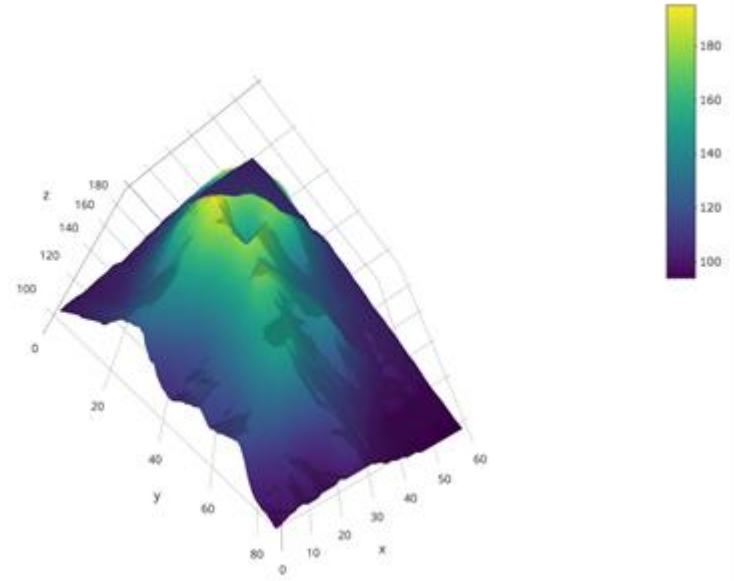
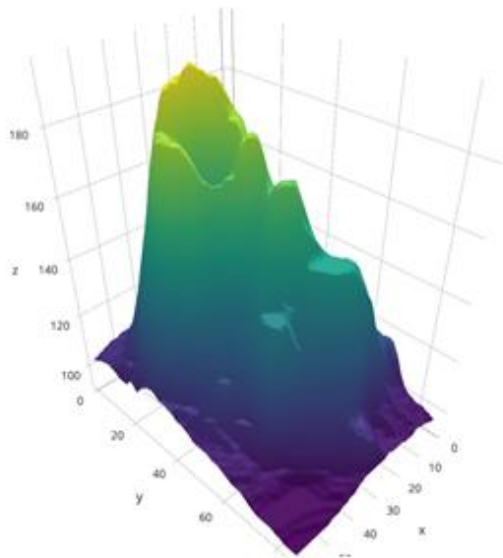
4. `plot <- plotly(data,.....)`
`api_create(plot, "...Doc path in plotly account")`

or

`plotly_POST()`

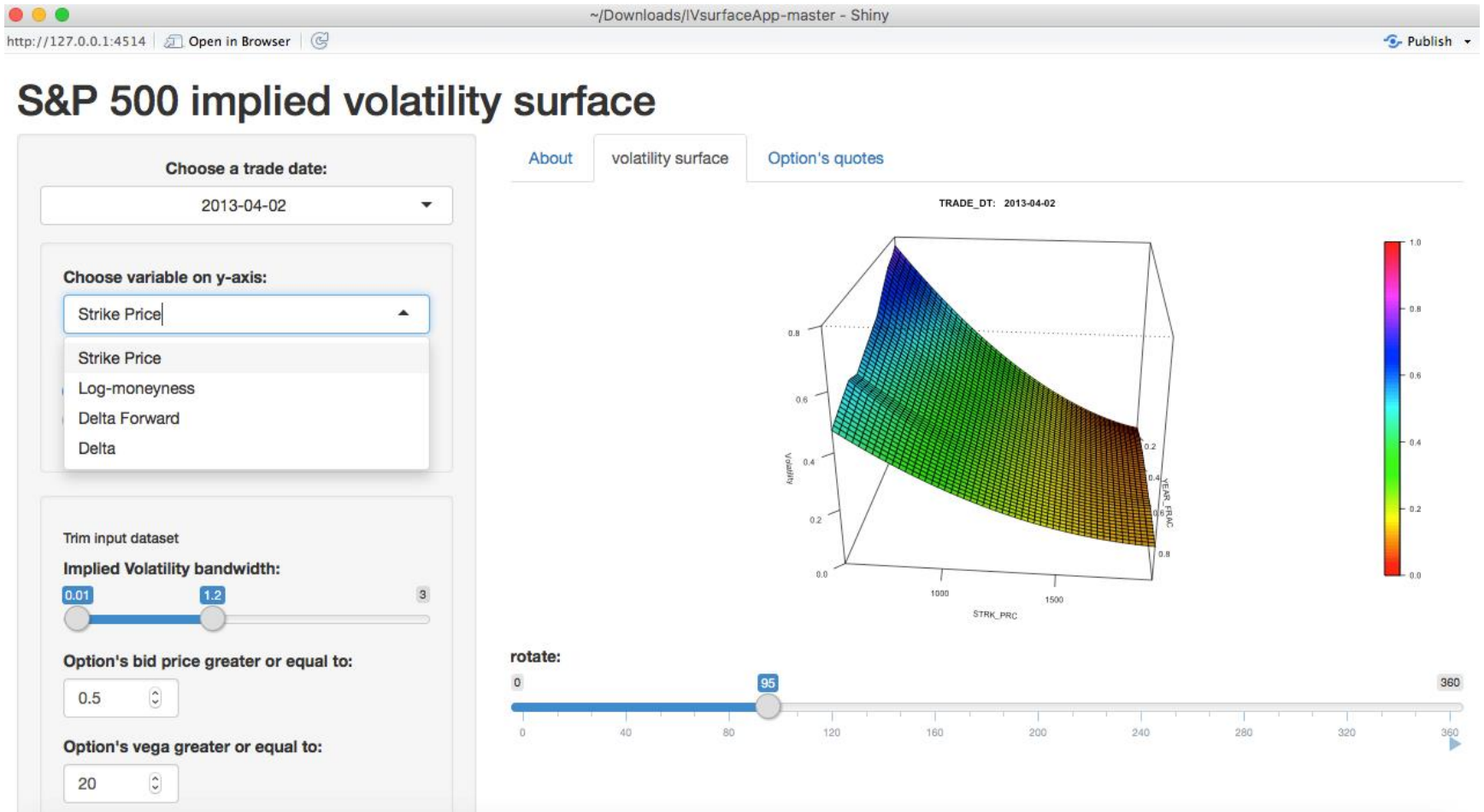


Interaction in Plotly





Shiny-Volatility Term Structure



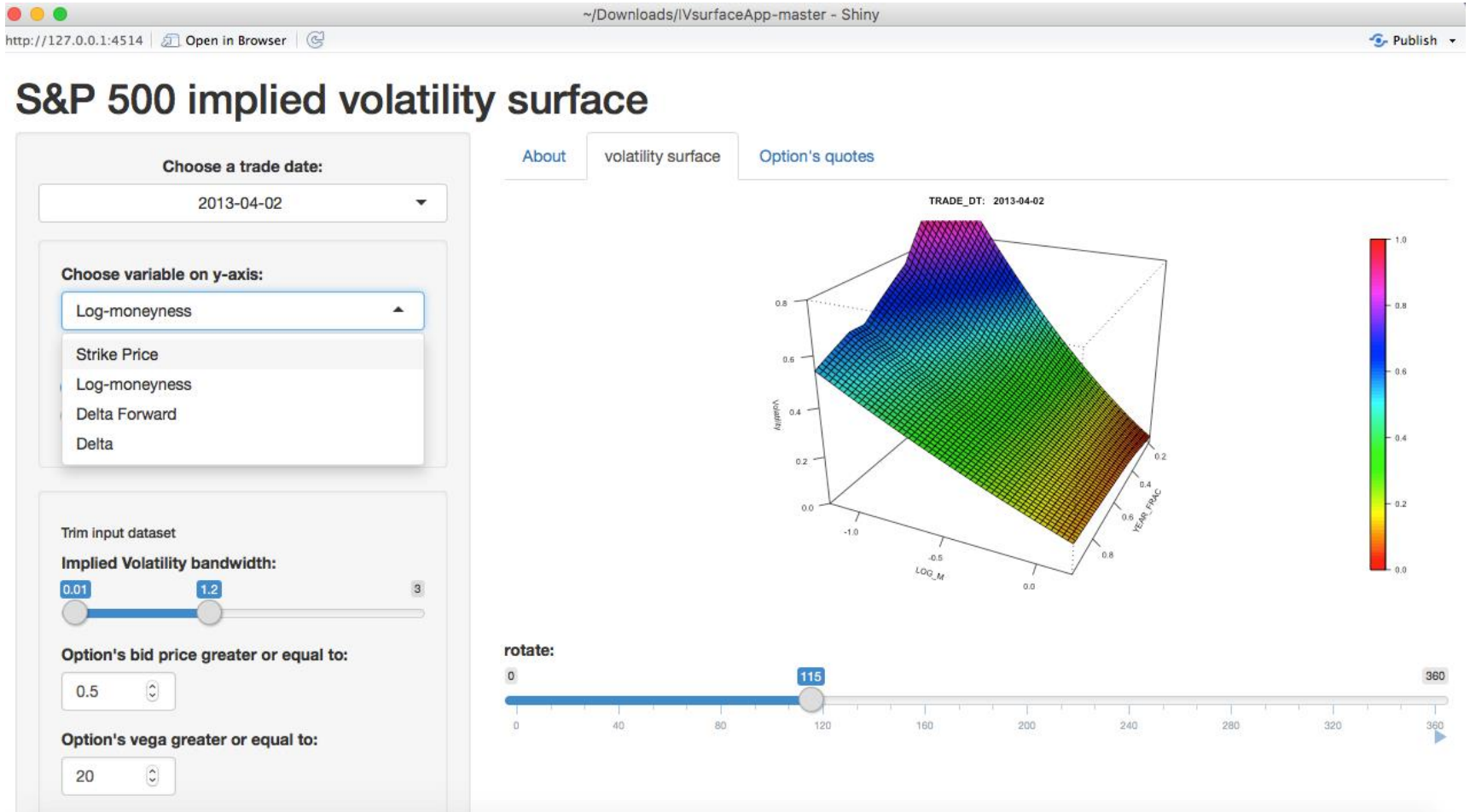


Shiny-UI Structure

- `shinyUI(pageWithSidebar(....`
- `wellPanel(`
 `selectInput("y.axis", "Choose variable on y-axis:",`
 `choices = c("Strike Price" = "STRK_PRC", "Log-`
 `moneyness" = "LOG_M", "Delta Forward" =`
 `"DELTA_FW", "Delta" = "DELTA"))`
- `mainPanel((.....(tabPanel("volatility surface",`
 `plotOutput("fig"),`
 `sliderInput("theta", "rotate:", min = 0L, max =`
 `360L, value = 45L, step = 5L, width = "100%")`



Shiny-Volatility Term Structure





Wrap up

- Don't throw away Excel...
- ...but use with caution
- R can be a valuable tool...
- ...for presenting, communicating and analysing
- Requires some investment of time/resources but it can offer an early payback



Questions

- Questions / comments / observations
- Suggestions

Pedro Ecija Serrano – pedro_ecija@yahoo.es

Conor Byrne – conor.byrne@permanenttsb.ie

Sean Cahill – sean.cahill@newireland.ie



Next presentation
