

VISUALIZATION OF AIRLINE SAFETY USING R

CSE3020 – DATA VISUALISATION

PROJECT-BASED COMPONENT REPORT

By

ROSHAN JULIUS (20BCB0001)

SURAJ RANGANATH (20BCB0060)

SHIKHAR NASHINE (20BCB0095)

DIVYANSH SHARMA (20BCT0138)

SHIVALIKA SINGH (20BCE2072)

School of Computer Science and Engineering



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

April 2022

DECLARATION

I hereby declare that the report entitled “VISUALISATION OF AIRLINE SAFETY USING R**” submitted by me, for the CSE3020 DATA VISUALISATION (EPJ) to VIT is a record of bonafide work carried out by me under the supervision of Dr.S.VENGADESWARAN .**

I further declare that the work reported in this report has not been submitted and will not be submitted, either in part or in full, for any other courses in this institute or any other institute or university.

Place : Vellore

Date : 28/04/2022

Signature of the Candidate

CONTENTS

	P.No
1. ABSTRACT	4
2. INTRODUCTION TO THE PROJECT	5
• OBJECTIVE	
• PROBLEM STATEMENT	
• FUNCTIONAL REQUIREMENTS	
3. DATA ABSTRACTION	6
4. TASK ABSTRACTION	11
5. DASHBOARD IMPLEMENTATION	18
6. RESULT ANALYSIS	26
7. CONCLUSION	28
8. APPENDIX	
• SCREEN SHOTS	29
• SAMPLE CODING	

1. ABSTRACT

Should Travellers Avoid Flying Airlines That Have Had Crashes in the Past? Is this behaviour rational? Are we less likely to fly in airlines that have had fatal crashes in the past, even if the crashes appear to be unrelated to them? Or are plane crashes really random events that happen at a similar rate on all airlines across time? Based on our airline-safety dataset, we can investigate this by looking at safety records for major commercial airlines over the last 30 years. The procedure is straightforward. The 30-year period will be divided into two halves: 1985-1999 and 2000-2014. Then we will look to see whether there was a correlation in crash rates from one half of the data set to the other. If we identify a correlation, that will imply that crash risk is persistent — predictable to some extent based on the airline

2. INTRODUCTION TO THE PROJECT

OBJECTIVE:

To visualise the statistics of fatal accidents and fatalities of different airlines during the time span of 1985-1999 and 2000-2014 to predict the success of these airlines in the coming years.

PROBLEM STATEMENT:

In today's generation we can see that there is an increase in the number of aircraft accidents all over the world, causing immense loss of life and money, thus the need to improve the safety of these aircrafts is of utmost priority.

With the help of the data set given and implementation of various visualisations, we can infer results that would help reduce such accidents and improve precautionary measures.

The main purpose is to examine and analyse the problems affecting aviation safety in the world and finding ways of mitigating them.

FUNCTIONAL REQUIREMENTS:

- 1) Data: Airline safety dataset - contains information about fatalities in different airlines from 1985-2014
- 2) Visualisation: An effective way to reduce the complexity of data and increase the simplicity and interactiveness.
- 3) Number Crunching: To point out the data that stands out and can help the user reach a conclusion.

3. DATA ABSTRACTION:

The dataset used for the project can be accessed using the below link:

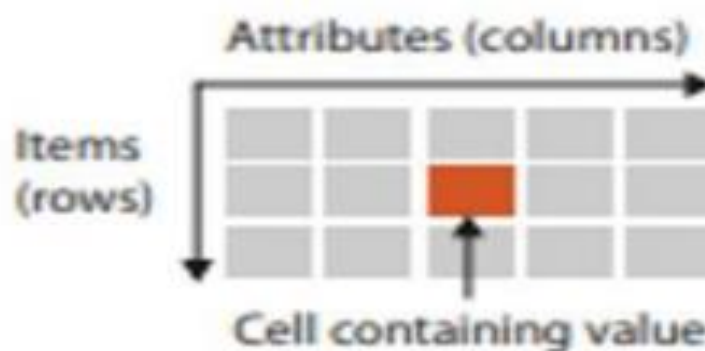
https://drive.google.com/file/d/1_KE1MTnj_rfYREiQkcGUI57DNLgI2ZSq/view

airline-safety.csv										
Open with ▾										
	A	B	C	D	E	F	G	H	I	J
1	airline	avail_seat_km_per_v	incidents_85_99	fatal_accidents_85_99	fatalities_85_99	incidents_00_14	fatal_accidents_00_14	fatalities_00_14	size	type
2	Aer Lingus	320906734	2	0	0	0	0	0	small	airbus
3	Aeroflot*	1197672318	76	14	128	6	1	88	small	airbus
4	Aerolineas Argentine	385803648	6	0	0	1	0	0	medium	airbus
5	Aeromexico*	596871813	3	1	64	5	0	0	medium	airbus
6	Air Canada	1865253802	2	0	0	2	0	0	large	boeing
7	Air France	3004002681	14	4	79	6	2	337	medium	boeing
8	Air India*	869253552	2	1	329	4	1	158	large	airbus
9	Air New Zealand*	710174817	3	0	0	5	1	7	medium	boeing
10	Alaska Airlines*	965346773	5	0	0	5	1	88	small	airbus
11	Alitalia	698012498	7	2	50	4	0	0	medium	boeing
12	All Nippon Airways	1841234177	3	1	1	7	0	0	large	airbus
13	American*	5228357340	21	5	101	17	3	416	small	boeing
14	Austrian Airlines	358239823	1	0	0	1	0	0	medium	airbus
15	Avianca	396922563	5	3	323	0	0	0	large	boeing
16	British Airways*	3179760952	4	0	0	6	0	0	small	airbus
17	Cathay Pacific*	2582459303	0	0	0	2	0	0	medium	boeing
18	China Airlines	813216487	12	6	535	2	1	225	small	airbus
19	Condor	417982610	2	1	16	0	0	0	medium	boeing
20	COPA	550491507	3	1	47	0	0	0	large	boeing
21	Delta / Northwest*	6525658894	24	12	407	24	2	51	small	airbus
22	Egyptair	557699891	8	3	282	4	1	14	medium	boeing
23	El Al	335448023	1	1	4	1	0	0	large	boeing

Data Set Description:

The dataset used for the project presents us the statistics of the number of fatalities and incidents that occurred with an airline during 2 different time periods. The 2 time periods that the dataset uses are from 1985 - 1999 and later 2000 - 2014. It provides us with the reports from 56 different airlines which operate across the globe.

- Dataset used: airline-safety.csv
- The above dataset is encoded in a csv format.
- The dataset is a static dataset which is of the Dataset type Tables
- The data types present in the following dataset are Items and Attributes.
- The dataset consists of 10 columns (Attributes) and 56 rows (Items)



Dataset Attributes: The description of the attributes present in the dataset used are listed below.

S.No	Attribute Name	Description
1.	Airline	The name of the airline
2.	Avail_seat_km_per_week	Kilometres travelled in one week
3.	Incidents_85_99	Number of incidences such as emergency landings during 1985- 1999
4.	Fatal_accidents_85_99	Number of fatal accidents during 198-1999
5.	Fatalities_85_99	Number of fatalities during 1985-1999
6.	Incidents_00_14	Number of incidences such as emergency landings during 2000-2014
7.	Fatal_accidents_00_14	Number of fatal accidents during 1985-1999
8.	Fatalities_00_14	Number of fatalities during 1985-1999
9.	Size	The size of the aircraft
10.	Type	The type of aircraft

The following table presents the attribute type present in the dataset and whether they are quantitative or categorical.

S.No	Attribute Name	Data Type
1.	Airline	Categorical
2.	Avail_seat_km_per_week	Quantitative (Discrete)
3.	Incidents_85_99	Quantitative (Discrete)
4.	Fatal_accidents_85_99	Quantitative (Discrete)
5.	Fatalities_85_99	Quantitative (Discrete)
6.	Incidents_00_14	Quantitative (Discrete)
7.	Fatal_Accidents_00_14	Quantitative (Discrete)
8.	Fatalities_00_14	Quantitative (Discrete)
9.	Size	Categorical
10.	Type	Categorical

Data Classification using Levels of Measurement.

S.No	Attribute Name	L.O.M
1.	Airline	Nominal
2.	Avail_seat_km_per_week	Ratio
3.	Incidents_85_99	Ratio
4.	Fatal_accidents_85_99	Ratio
5.	Fatalities_85_99	Ratio

6.	Incidents_00_14	Ratio
7.	Fatal_Accidents_00_14	Ratio
8.	Fatalities_00_14	Ratio
9.	Size	Nominal
10.	Type	Nominal

It can be inferred that the dataset consists of:

3 Nominal Attribute Variables:

These attributes are being used as labels and categories that allow us to refer to which airline the fatalities or accidents have occurred more upon. The data values in particular don't have any order associated with them. The Size and Type attribute provide the extra information associated with the aircraft, whether the aircraft was of type Boeing or Airbus and the relative size being small, medium, or large to accommodate the given passengers travelling.

7 Ratio Attribute Variables:

These values indicate the number of fatalities or incidents that occurred with an airline service corresponding to the given 2 time periods.

A wide range of values can be seen in the dataset providing us the necessary information to highlight the trend.

It can also be inferred the attribute values are sequential as they range from a minimum (0) to maximum.

4. TASK ABSTRACTION:

Aim: -To analyse the future of airline business by visualising the past accidents and fatalities in order to determine the success of corresponding Airline.

Actions

Analyse:

Consume:

Discover: To analyse the trend of the success of the airlines the airlines given in the dataset were plotted against the size of the aeroplane to observe whether the number of accidents that were occurring had any relation to the size of the aircraft that was being operated. To analyse the dataset based on the 2 different time periods and find the airlines with the highest fatality rate amongst them.

Present: To present the data various idioms such as stacked bar graph, line chart, pie chart have been implemented for better visualisation of the dataset according to the user choice. You can toggle between the 2 different Time periods to compare the airline statistics.

Produce:

Annotate: Through the given visualisation we can find the exact statistical value of the fatality rate of a given airline in the given time period.

Derive: Fatality per km travelled, normalises the distance travelled for every airline thus providing a level footing to compare fatalities of different airlines.

Search:

Lookup: The airline with the most number of fatalities can be searched by observing the graphs with the highest plot. The different visualisation idioms allow you to easily perceive the airline with the most fatality rate.

Browse: Through the given virtualisation you can browse the different airlines and observe the fatalities rate that has occurred for an individual airline in the given 2 time periods.

Query:

Identify: Visualisations have been made to observe the type of aircraft being used with respect to the size of the aircraft. This allows us to identify if the type of aircraft being used and size of the aircraft are compatible for flight or not.

Compare: The aim was to analyse the past data and current data of the airline fatality count and determine whether the airline has worked on its service to improve the following statistics.

Targets

All Data:

Trend:

The fatalities that had occurred during the 2 time periods can be observed and the trend for the fatality rate can be seen whether the given airline continues to suffer more accidents. By observing this the user can infer whether the airline has worked towards improving the quality of the aircraft to prevent such mishaps to occur in the coming future by observing the trend if it increases or decreases.

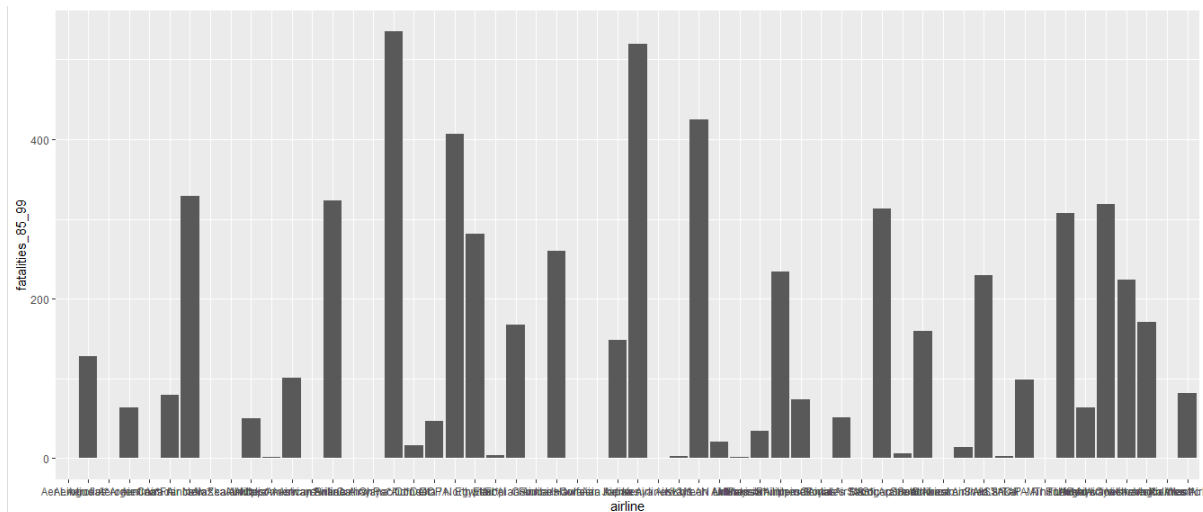
Attributes:

The number of fatalities that have occurred can be determined according to the time period or according to the airline to find the correlation between the fatalities rate in the given 2 time periods. It also allows the user to determine the fatality count of an individual airline with respect to that of other airlines. This helps to determine the safest airline to travel.

Questions and Answers [identifying Actions and Targets]: -

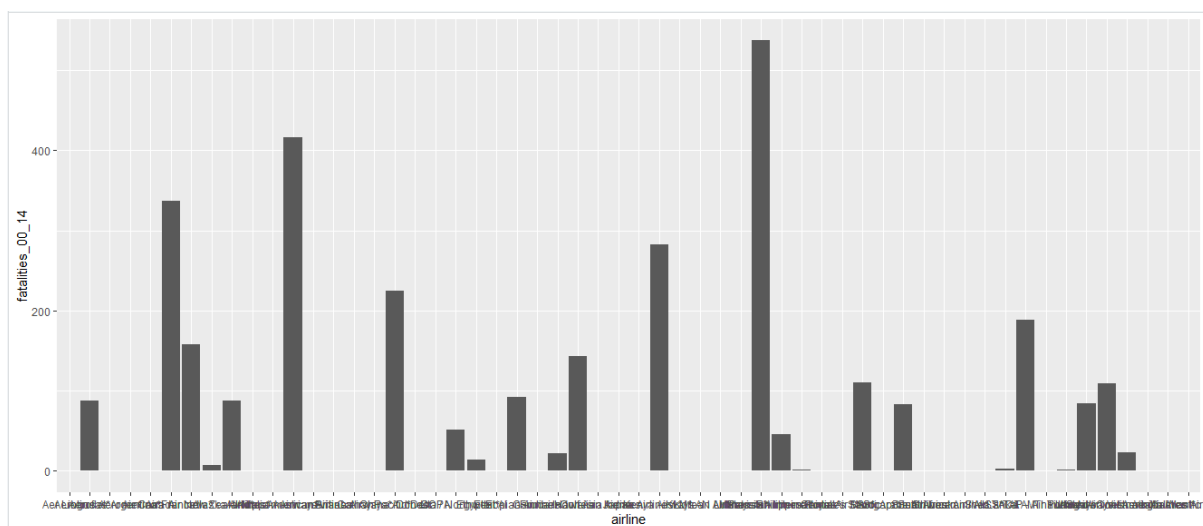
Q1) Which of the airlines had faced more than 400 fatalities during 1985-1999?

Ans) Attributes required: airline, fatalities_85_99



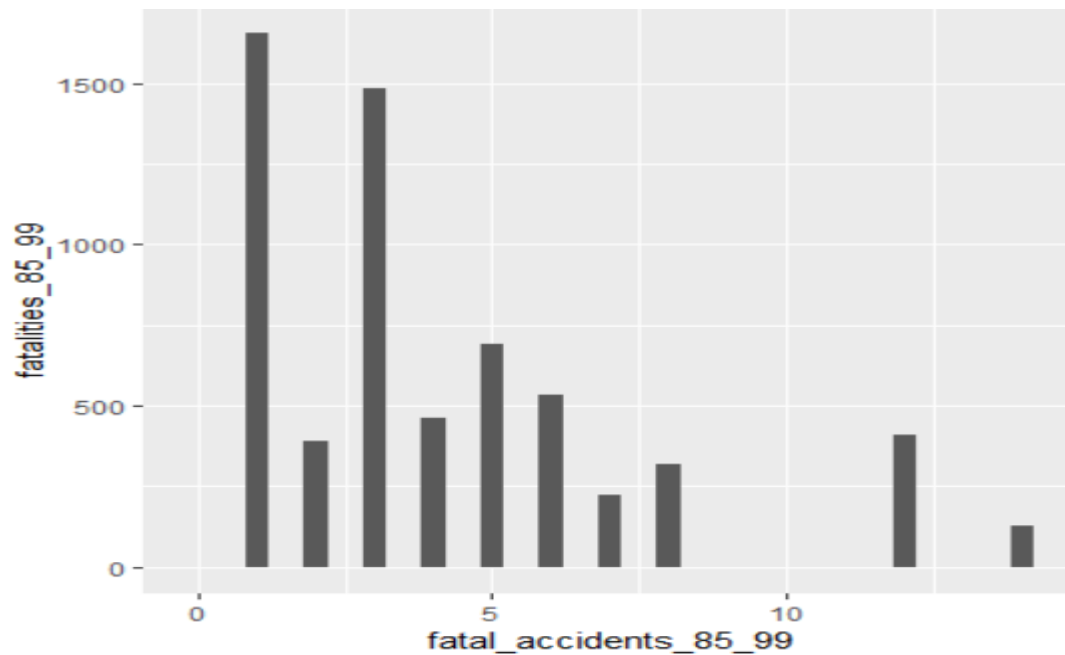
Q2) Which of the airlines had faced more than 400 fatalities during 2000-2014?

Ans) Attributes required: airline, fatalities_00_14



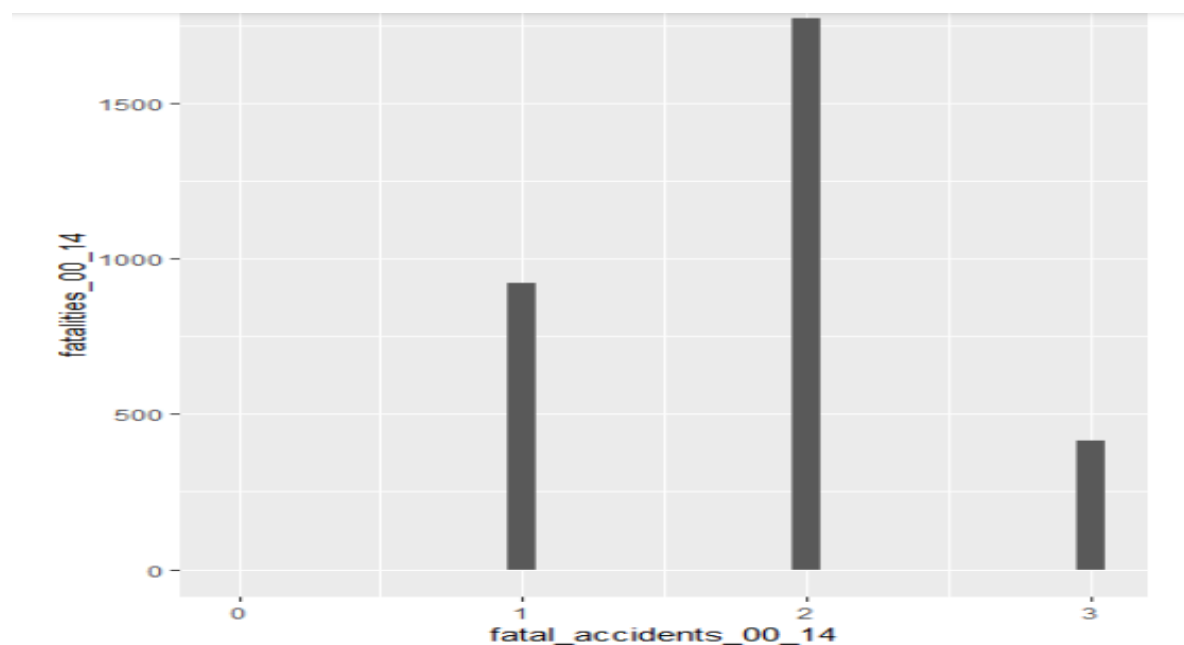
Q3) How many fatal accidents led to fatalities of these airlines during 1985-1999?

Ans) Attributes required: fatal_accidents_85_99, fatalities_85_99



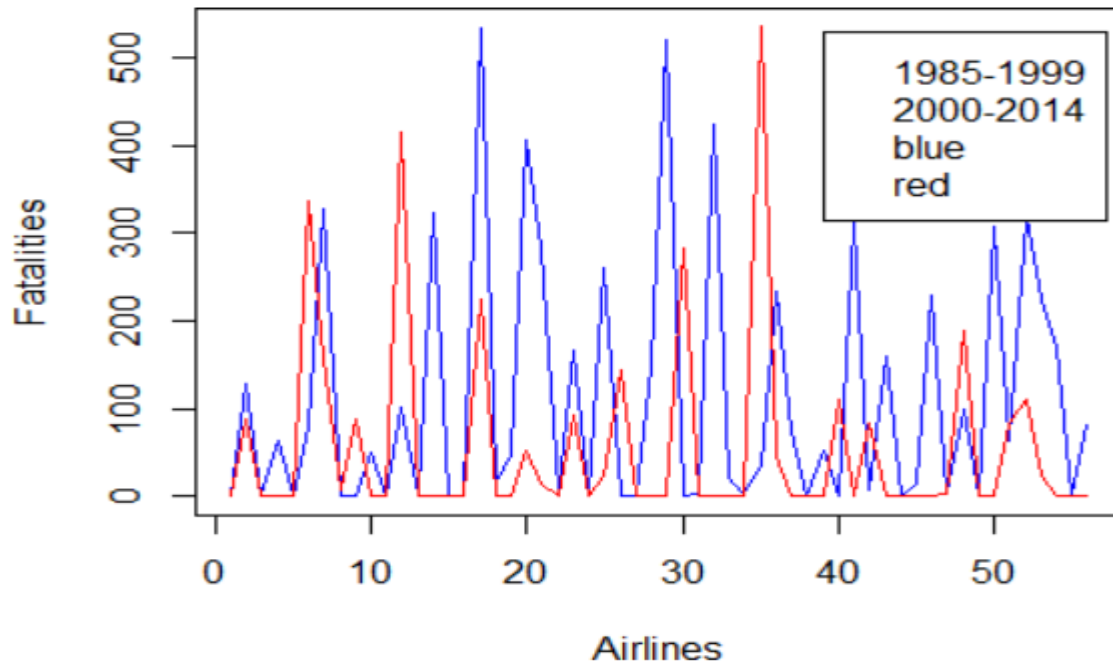
Q4) How many fatal accidents led to fatalities of these airlines during 2000-2014?

Ans) Attributes required: fatal_accidents_00_14, fatalities_00_14



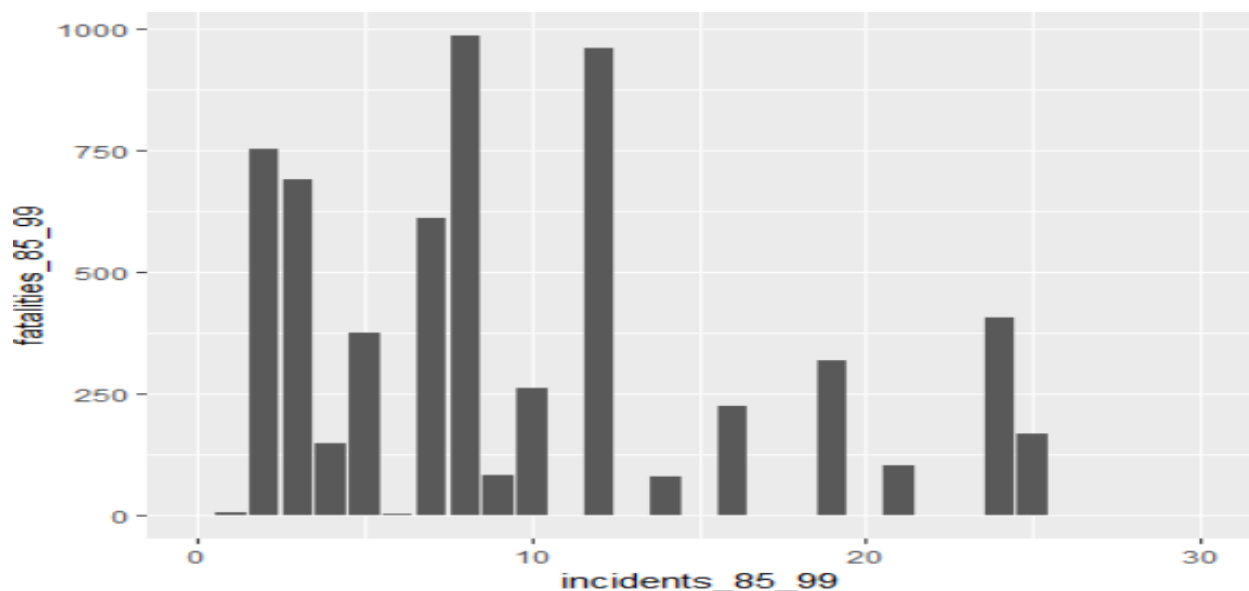
Q5) Comparing the fatalities of 1985-1999 and 2000-2014, has the number of fatalities decreased or increased?

Ans) Attributes required: fatalities_85_99, fatalitites_00_14



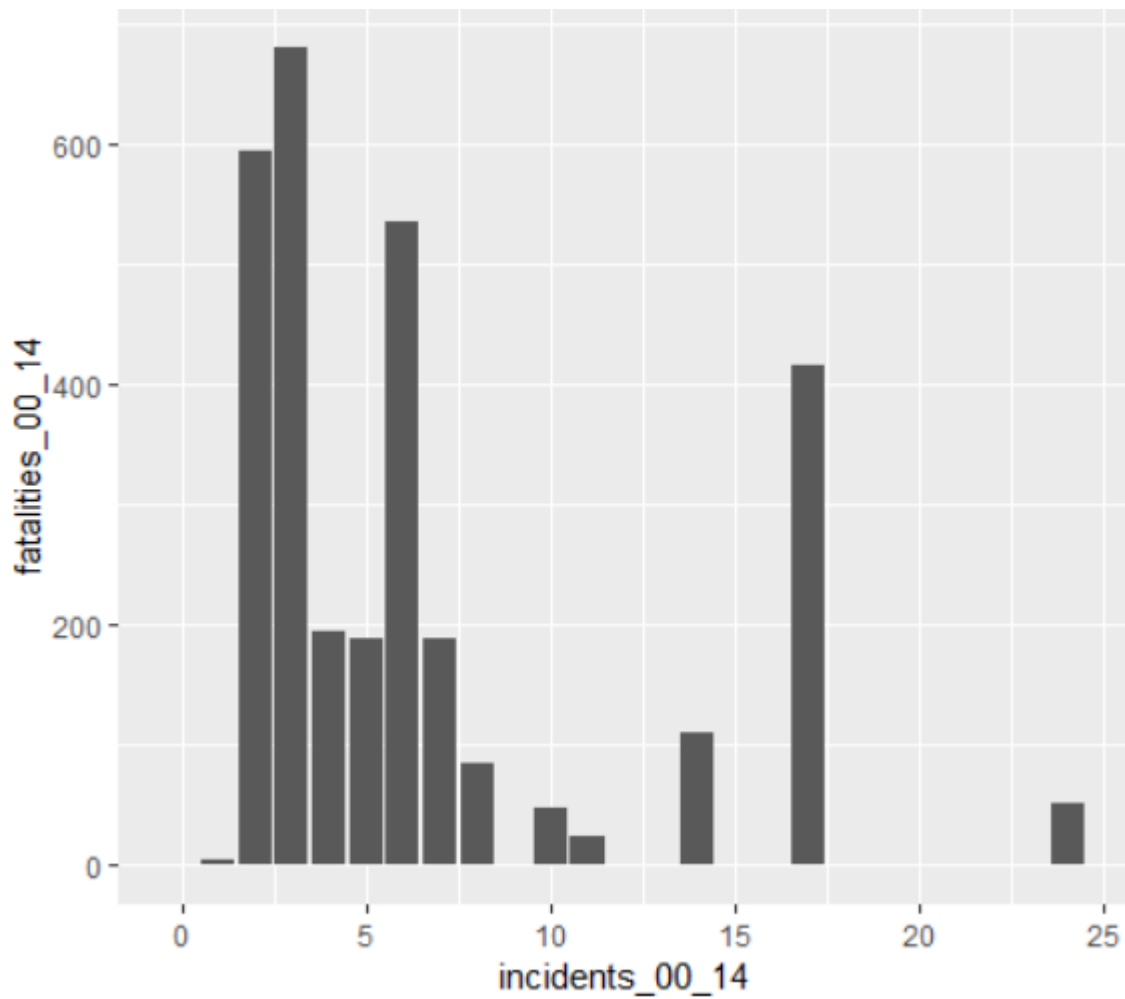
Q6) How many incidents have led to fatal accidents of these airlines during 1985-1999?

Ans) Attributes required: incidents_85_99, fatalities_85_99



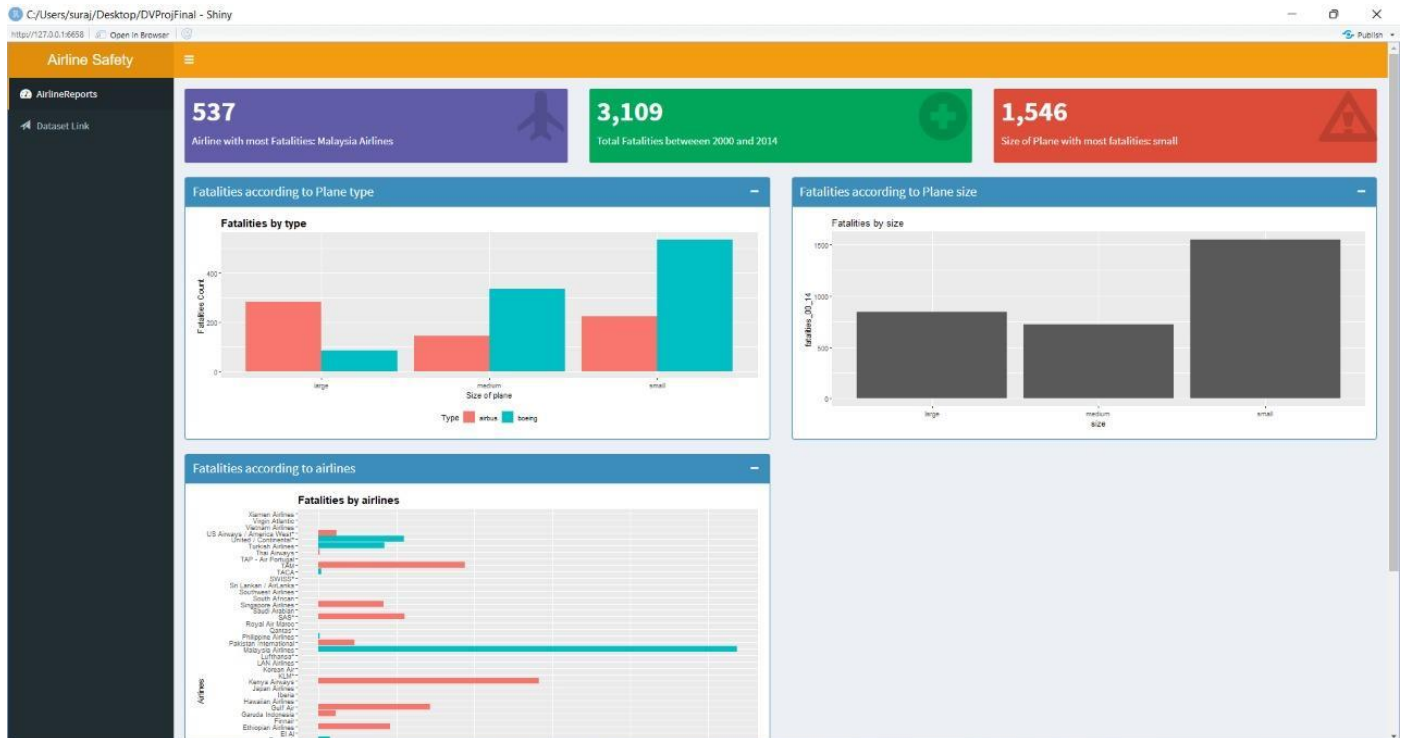
Q7) How many incidents have led to fatal accidents of these airlines during 2000-2014?

Ans) Attributes required: incidents_00_14, fatalities_00_14



5. DASHBOARD IMPLEMENTATION

Implementing the code on R shiny:



```
library(shiny)
```

```
require(shinydashboard)
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
airline <- read.csv("airline-safety.csv",stringsAsFactors =  
F,header=T)
```

```
header <- dashboardHeader(title = "Airline Safety Dashboard")
```

```
sidebar<-dashboardSidebar(sidebarMenu(  

```

```
menuItem(" AirlineReports",tabName="dashboard2",icon=icon("dash  
board", verify_fa = FALSE)),
```

```
menuItem("Dataset Link", icon = icon("send",lib='glyphicon'), href =  
"https://drive.google.com/file/d/1_KE1MTnj_rfYREiQkcGUI57DNL  
gI2ZSq/view?usp=sharing"))))
```

```
frow1 <- fluidRow( valueBoxOutput("value1")  
,valueBoxOutput("value2") ,valueBoxOutput("value3"))
```

```
frow2<-fluidRow(box( title = "Fatalities according to Plane  
type",status = "primary" ,solidHeader = TRUE , collapsible = TRUE,  
plotOutput("fatalitiesbytype", height = "300px") ),box( title =  
"Fatalities according to Plane size",status = "primary",solidHeader =  
TRUE,collapsible = TRUE ,plotOutput("fatalitiesbysize", height =  
"300px") ) ,box( title = "Fatalities according to airlines" ,status =  
"primary" ,solidHeader = TRUE ,collapsible = TRUE  
,plotOutput("fatalitiesbyairline", height = "600px") ))
```

```
body <- dashboardBody(frow1, frow2)
```

```
ui <- dashboardPage(title = 'Airline Safety ', header, sidebar, body,  
skin='yellow')
```

```
server <- function(input, output) {
```

```
  total.fatalities <- sum(airline$fatalities_00_14)
```

```
  total.airlines <- airline %>% group_by(airline) %>%  
summarise(value = sum(fatalities_00_14)) %>%  
filter(value==max(value))
```

```
  total.size <- airline %>% group_by(size) %>% summarise(value =  
sum(fatalities_00_14)) %>% filter(value==max(value))
```

```
  output$value1<-renderValueBox({valueBox(  
formatC(total.airlines$value, format="d", big.mark=','),paste('Airline  
with most Fatalities:',total.airlines$airline) ,icon =  
icon("plane",lib='glyphicon') ,color = "purple"))})
```

```
output$value2<-renderValueBox({ valueBox(formatC(total.fatalities,
format="d", big.mark=','), 'Total Fatalities between 2000 and
2014',icon = icon("plus-sign",lib='glypicon') ,color = "green"))})
```

```
output$value3<-
renderValueBox({ valueBox(formatC(total.size$value, format="d",
big.mark=','), paste('Size of Plane with most fatalities:',total.size$size)
,icon = icon("warning-sign",lib='glypicon') ,color = "red"))})
```

```
output$fatalitiesbytype<-renderPlot({ ggplot(data=airline, aes(x=size,
y=fatalities_00_14, fill=factor(type))) + geom_bar(position =
"dodge", stat = "identity") + ylab("Fatalities Count") + xlab("Size of
plane") + theme(legend.position="bottom"
,plot.title = element_text(size=15, face="bold")) + ggtitle("Fatalities
by type") + labs(fill = "Type"))})
```

```
output$fatalitiesbysize<-renderPlot({ ggplot(data=airline, aes(x=size,
y=fatalities_00_14)) + geom_bar(position = "stack", stat = "identity")
+ ggtitle("Fatalities by size") + labs(fill = "Type"))})
```

```
output$fatalitiesbyairline<-
renderPlot({ ggplot(data=airline,aes(x=airline, y=fatalities_00_14,
fill=factor(type))) + geom_bar(position = "dodge", stat = "identity") +
ylab("Fatalities") + xlab("Airlines") +
theme(legend.position="bottom" ,plot.title = element_text(size=15,
face="bold")) + ggtitle("Fatalities by airlines") + labs(fill = "Type") +
coord_flip())})
```

```
shinyApp(ui,server)
```

Implementing the code on Flexdashboard:

title: "Visualization of Airlines Fatalities"

output:

```
flexdashboard::flex_dashboard:
```

```
  orientation: columns
```

```
  vertical_layout: fill
```

```
---
```

```
{r setup, include=FALSE}
```

```
library(flexdashboard)
```

```
airline <- read.csv("airline-safety.csv",stringsAsFactors =  
F,header=T)
```

Graphs

```
=====
```

```
Column {data-width=500}
```

```
-----
```

```
### Chart A
```

```
{r}
```

```
library(plotly)
```

```
airline<-read.csv("airline-safety.csv")
```

```
p<-
```

```
plot_ly(airline,x=airline$fatalities_85_99,y=airline$airline,type='bar',  
orientation = 'h', name = 'Fatalities between 1985 to 1999',marker =  
list(color = 'rgba(246, 78, 139, 0.6)',line = list(color = 'rgba(246, 78,  
139, 1.0)',width = 3))) %>%
```

```
add_trace(x = airline$fatalities_00_14, name = 'Fatalities between 2  
000 to 2014',marker = list(color = 'rgba(58, 71, 80, 0.6)',line =  
list(color = 'rgba(58, 71, 80, 1.0)',width = 3))) %>%
```

```
layout(barmode = 'stack',xaxis = list(title = ""), yaxis = list(title = ""))
```

```
p
```

```
Column {data-width=500}
```

```
-----
```

```
### Chart B
```

```
{r}
```

```
plot(airline$fatalities_85_99,type="l",col="orange",xlab="Airlines",yl  
ab="Fatalities")
```

```
lines(airline$fatalities_00_14,type="l",col="green")
```

```
legend(39,530,legend<-c("1985-1999","2000-2014",col=c("orange","green"))))
```

Chart C

```
{r}
```

```
p <- plot_ly(airline, x = airline$airline, y = airline$fatalities_85_99,
name = 'Fatalities 1985-1999', type = 'scatter', mode = 'lines',
```

```
line = list(color = 'rgb(205, 12, 24)')) %>%
```

```
add_trace(y = airline$fatalities_00_14, name = 'Fatalities 2000-2014 ',
line = list(color = 'rgb(22, 96, 167)'))
```

```
p
```

Pie Charts

```
=====
```

Column {data-width=500}

```
-----
```

Chart A

```
{r}
```

```
data<-read.csv("airline-safety.csv")
```

```
p<-  
plot_ly(data,labels=~data$airline,values=~data$fatalities_85_99,type  
='pie')%>%
```

```
layout(title = 'Fatalities between 1985 and 1999', xaxis = list(showgrid  
= FALSE, zeroline = FALSE, showticklabels = FALSE), yaxis =  
list(showgrid = FALSE, zeroline = FALSE, showticklabels =  
FALSE))
```

p

Column {data-width=500}

Chart B

{r}

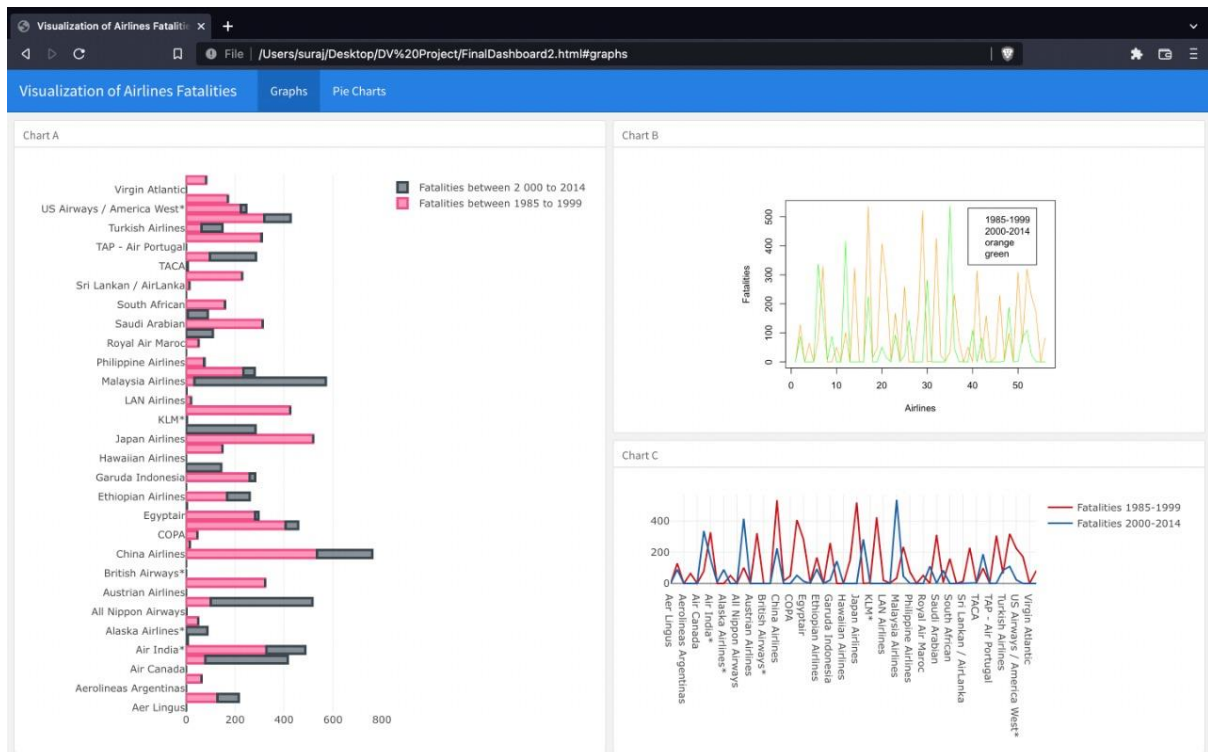
```
data<-read.csv("airline-safety.csv")
```

```
p<-  
plot_ly(data,labels=~data$airline,values=~data$fatalities_00_14,type  
='pie')%>%
```

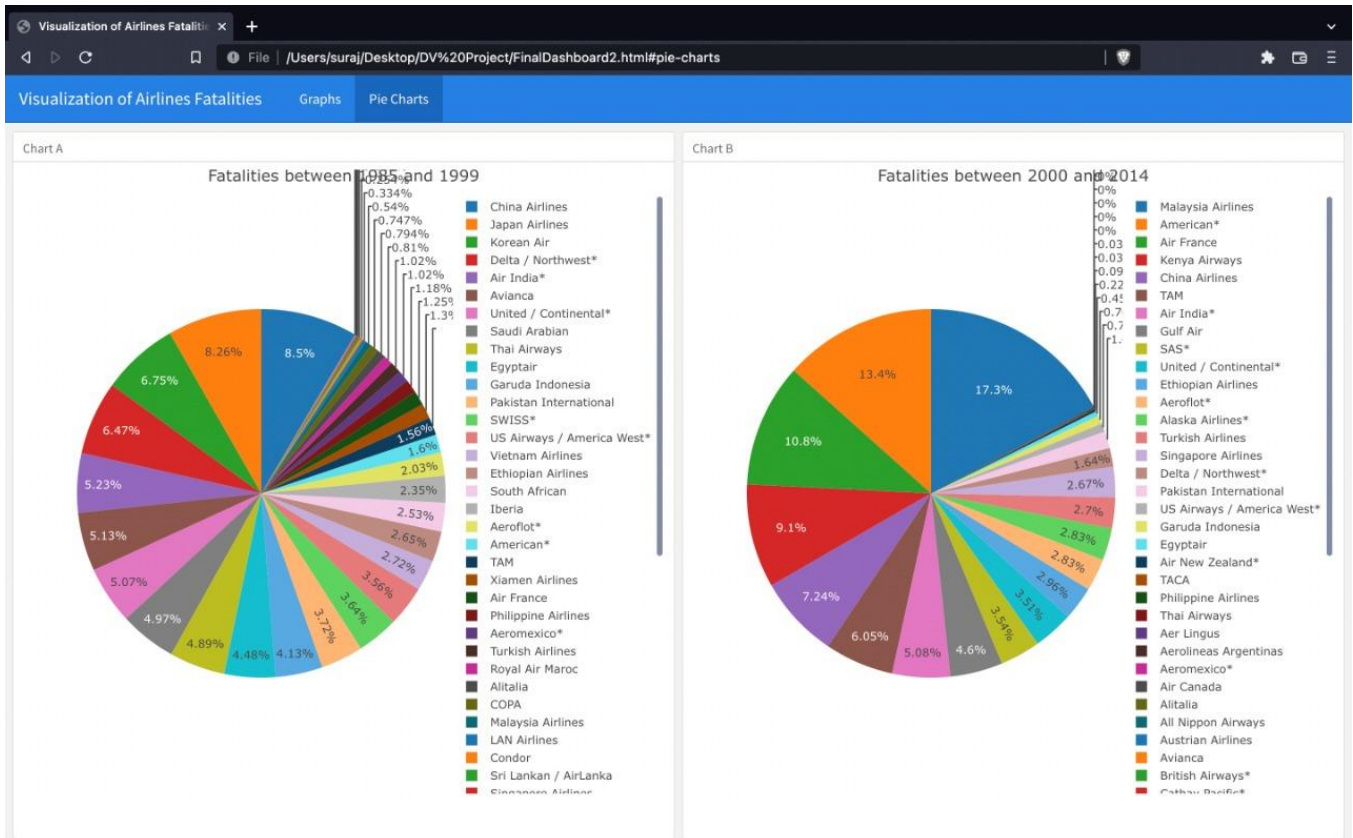
```
layout(title = 'Fatalities between 2000 and 2014', xaxis = list(showgrid  
= FALSE, zeroline = FALSE, showticklabels = FALSE), yaxis =  
list(showgrid = FALSE, zeroline = FALSE, showticklabels =  
FALSE))
```

p

- Graph:



● PieChart:



6. RESULT ANALYSIS

Interpretation for the above dashboards is given as follows

R - shiny dashboard:

1. Fatalities by type

We observe that in medium and large planes Boeing suffers more fatalities than airbus however in small planes, airbus suffers more fatalities than Boeing. This leads one to believe that Airbus is better at building reliable medium and large planes whereas Boeing is better at building reliable small planes.

2. Fatalities by size

We observe that small planes suffer higher fatalities than larger planes. This could be because smaller planes have lesser safety features than bigger planes and are more easily Sayed by change in air currents.

3. Fatalities by airlines

We observe that Malasian airlines have the highest fatality rate by a significant margin. This indicates that the airline has to step up its game to ensure the safety of its passengers.

R flexdashboard:

1. Graphs:

A. Chart A - Horizontal Stacked Bar Graph showing Fatalities of different airlines

The horizontal bar graph shows the fatalities of different airlines during two time periods stacked on top of each other. Overall, it can be concluded that China Airlines has had the most number of fatalities.

B. Chart B - Multiple line plot comparing the fatality per km travelled for every airline

This chart normalises the dist travelled for every airline, thus providing a level footing to compare fatalities of different airlines in both the time periods 2000-2014 and 1985-1990

C. Chart C - Multiple line plots comparing the fatalities of airline

Using multiple line plots in the same graph, we can conclude that the number of fatalities for most of the airlines have decreased. But some airlines have had more fatalities in 2000-2014 than during 1985-1999.

2. Pie chart:

A. Chart A - The above pie chart shows the fatalities in the year 1985-1999. The different colours make it easier to distinguish the different airlines. Japan Airlines show the most number of fatalities.

B. Chart B - The pie chart shows the fatalities in the year 2000-2014. The different colours make it easier to distinguish the different airlines. Malaysia Airlines show the most number of fatalities.

7.CONCLUSION

It can be seen there has a decrease in the number of fatalities from 6,295 to 3,109. So airlines have become more reliable. But some airlines still cause a lot of fatalities and their issues much be looked into to make travel feel safe for its customers.

8.APPENDIX

SCREENSHOTS AND SAMPLE CODING

The screenshot shows a web browser window with a Shiny dashboard titled "Airline Safety Dashboard". The dashboard is divided into a sidebar and a main content area. The sidebar contains a menu with "Airline Reports" selected. The main content area displays three plots: "Fatalities according to Plane type", "Fatalities according to Plane size", and "Fatalities according to airlines". The "Fatalities according to Plane type" plot shows a bar chart with "primary" as the only category. The "Fatalities according to Plane size" plot shows a bar chart with "primary" as the only category. The "Fatalities according to airlines" plot shows a bar chart with "primary" as the only category. The dashboard also includes a text output box for "Total Fatalities between 2000 and 2014" and a text output box for "Size of Plane with most fatalities".

```
1 ---
2 title: "Visualization of Airlines Fatalities"
3 output:
4   flexdashboard::flex_dashboard:
5     orientation: columns
6     vertical_layout: fill
7 ---
8
9 {r setup, include=FALSE}
10 library(flexdashboard)
11 airline <- read.csv("airline-safety.csv", stringsAsFactors = F, header=T)
12
13
14
15 Graphs
16
17 Column (data-width=500)
18
19
20 ## Chart A
21
22 {r}
23 library(plotly)
24 airline <- read.csv("airline-safety.csv")
25 p <- plot_ly(airline, x=airline$fatalities_85_99, y=airline$airline, type='bar',
26   orientation = 'h', name = 'Fatalities between 1985 to 1999',
27   marker = list(color = 'rgba(246, 78, 139, 0.6)'),
28   line = list(color = 'rgba(246, 78, 139, 1.0)',
29     width = 3)) %>%
30   add_trace(x = airline$fatalities_00_14, name = 'Fatalities between 2000 to 2014',
31     marker = list(color = 'rgba(58, 71, 80, 0.6)'),
32     line = list(color = 'rgba(58, 71, 80, 1.0)',
33       width = 3)) %>%
34   layout(barmode = 'stack',
35     xaxis = list(title = ""),
36     yaxis = list(title = ""))
37 p
38
39
40 Column (data-width=500)
41
42
43 ## Chart B
44
45 {r}
```

