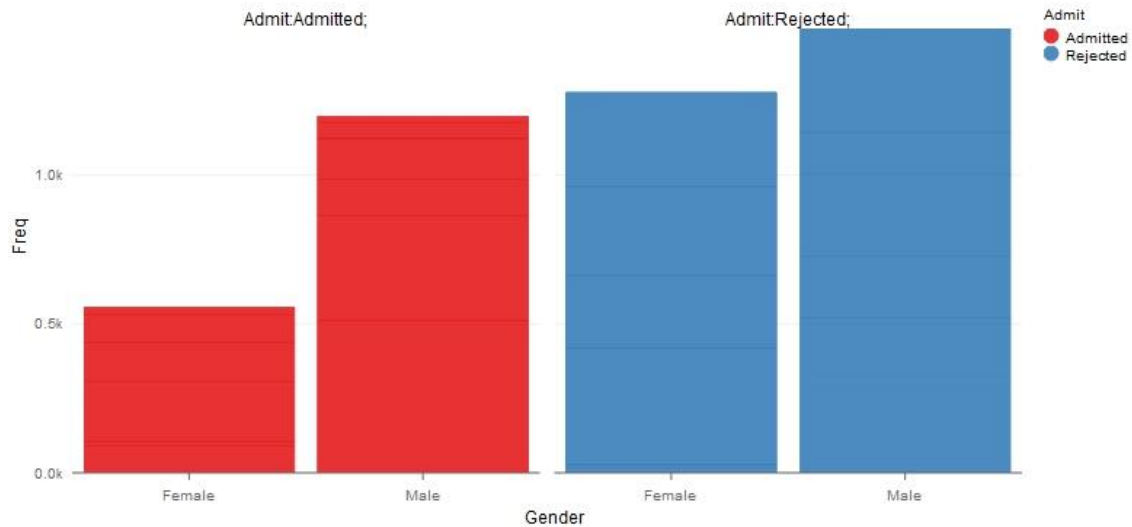


1- University of Arizona Task

Based on the data I found that there is a discrimination, and the results were as below:

	<i>Admitted</i>		<i>Rejected</i>	
<i>Female</i>	557	30.35%	1278	69.65%
<i>Male</i>	1198	44.52%	1493	55.48%

As we can see that female admission rate is 30.35 %, and male admission rate is 44.51%, so there is a discrimination.



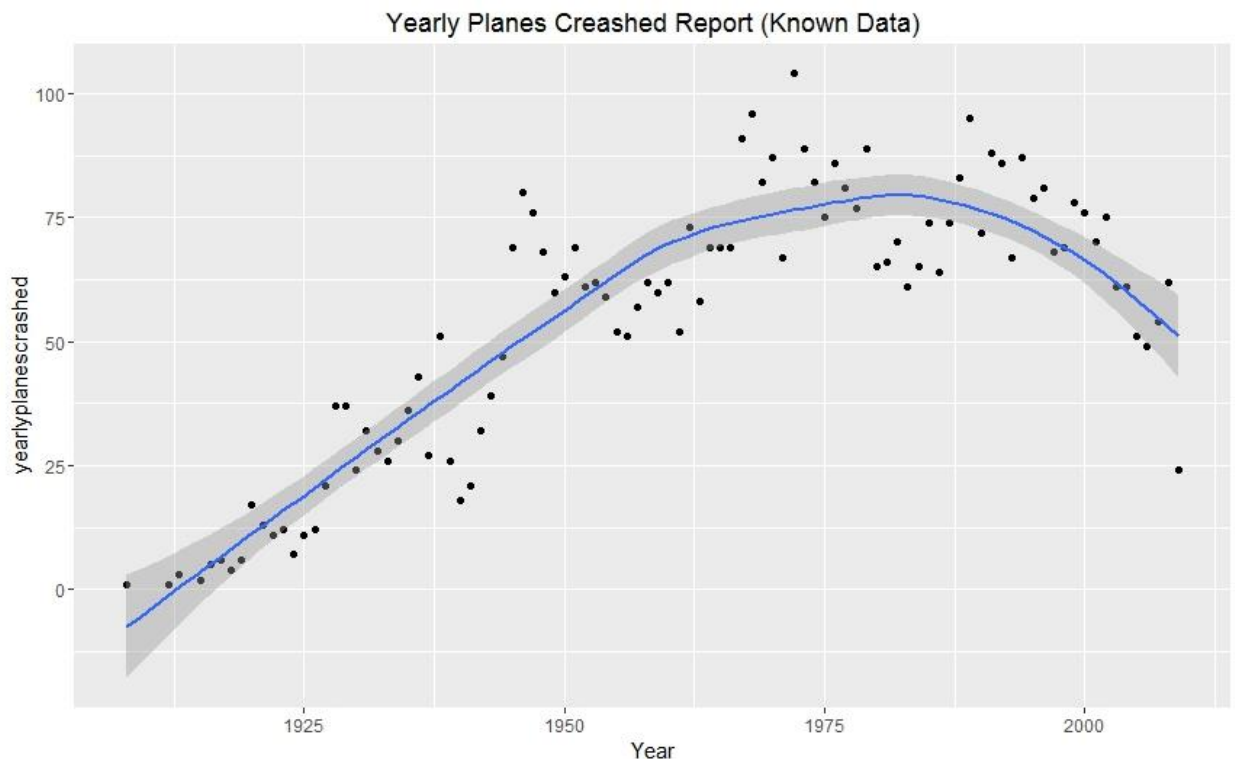
I've solved with 3 solutions which can be found in "**University.R**" file, but there are many other solutions that also can be used. Results have been extracted in "**University_Gender_Admissions.csv**", and "**University_Gender_Admissions_probabilities.csv**" files, and the chart has been extracted in "**University_Gender_Admissions.html**" file.

2- Air crash dataset

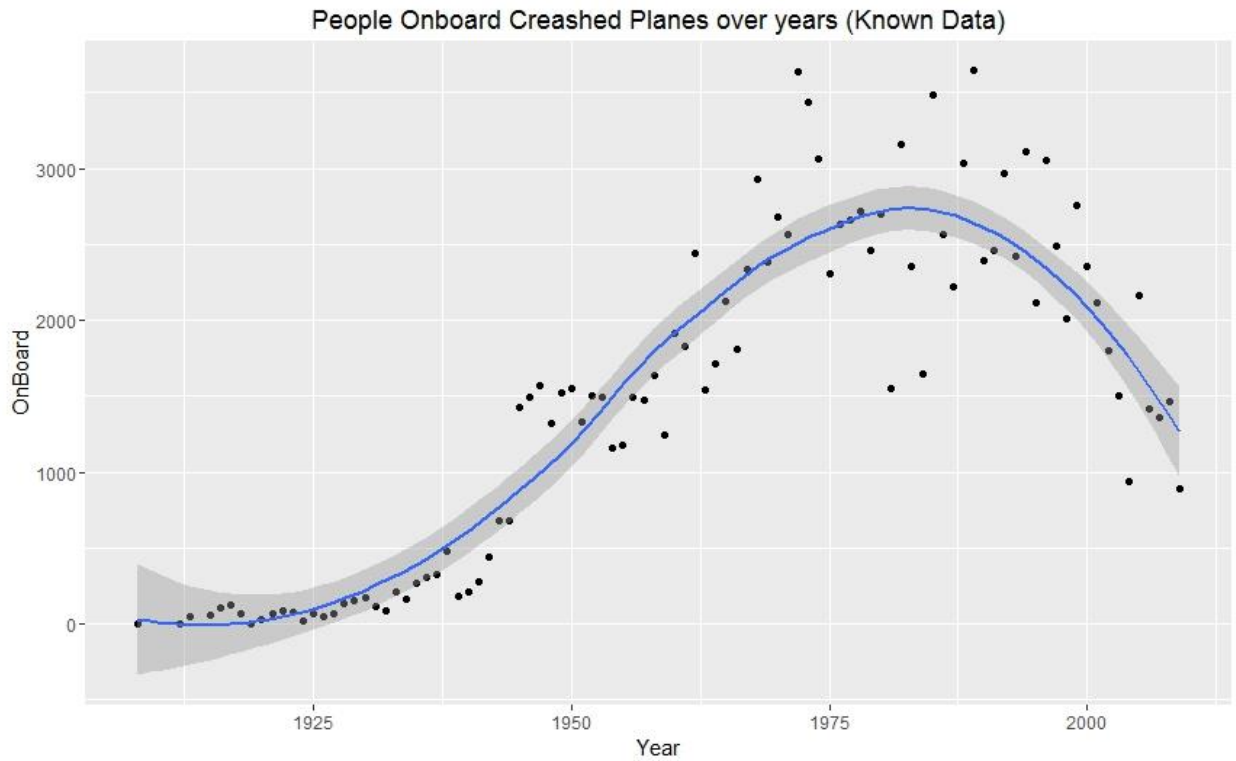
Solution for this task can be found in “**AirplaneCrashesSince1908.R**” file.

Just for info:

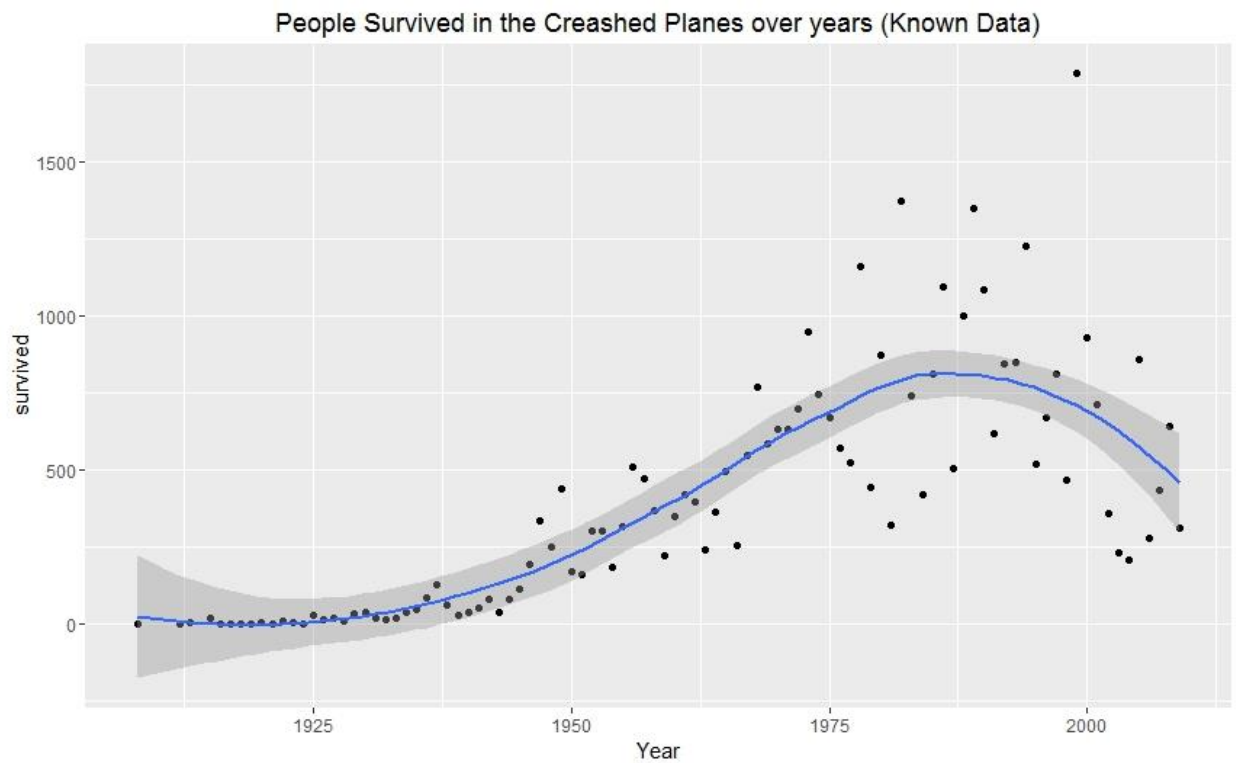
- Number of Crashed planes with Unknown number of people onboard: 22 ~ can be neglected because they are less than 0.5%
 - Number of Crashed planes with Unknown number of people died: 12 ~ can be neglected because they are less than 0.5%
 - Number of Crashed planes with Unknown number of people survived: 22 ~ can be neglected because they are less than 0.5%
- a- Yearly how many planes crashed? how many people were on board? how many survived? how many died?
- Result can be found in the “**YearlyPlanesCrashedReportKnown.csv**” file for the flights with all known data which are number of recorded people onboard or number of recorded people survived or died.



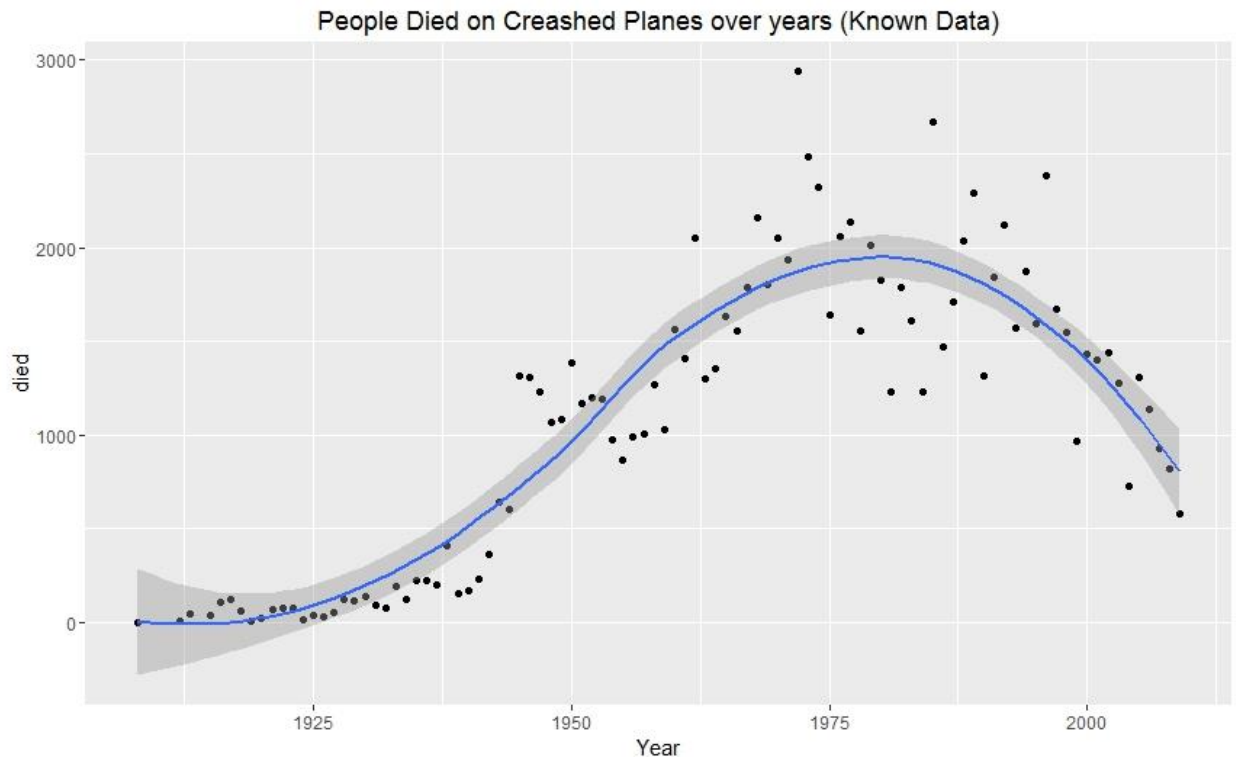
As noticed that the most number of plane crashes are in mid 70's, and the maximum recorded number of crashes happened was in 1972 with 104 planes crashed. Also we can notice that number of plane crashes started to decrease again after 90's, which means that flight researches are in the right direction to make safer flights.



We can notice that people onboard of these crashed planes was at the highest level in the mid 80's.

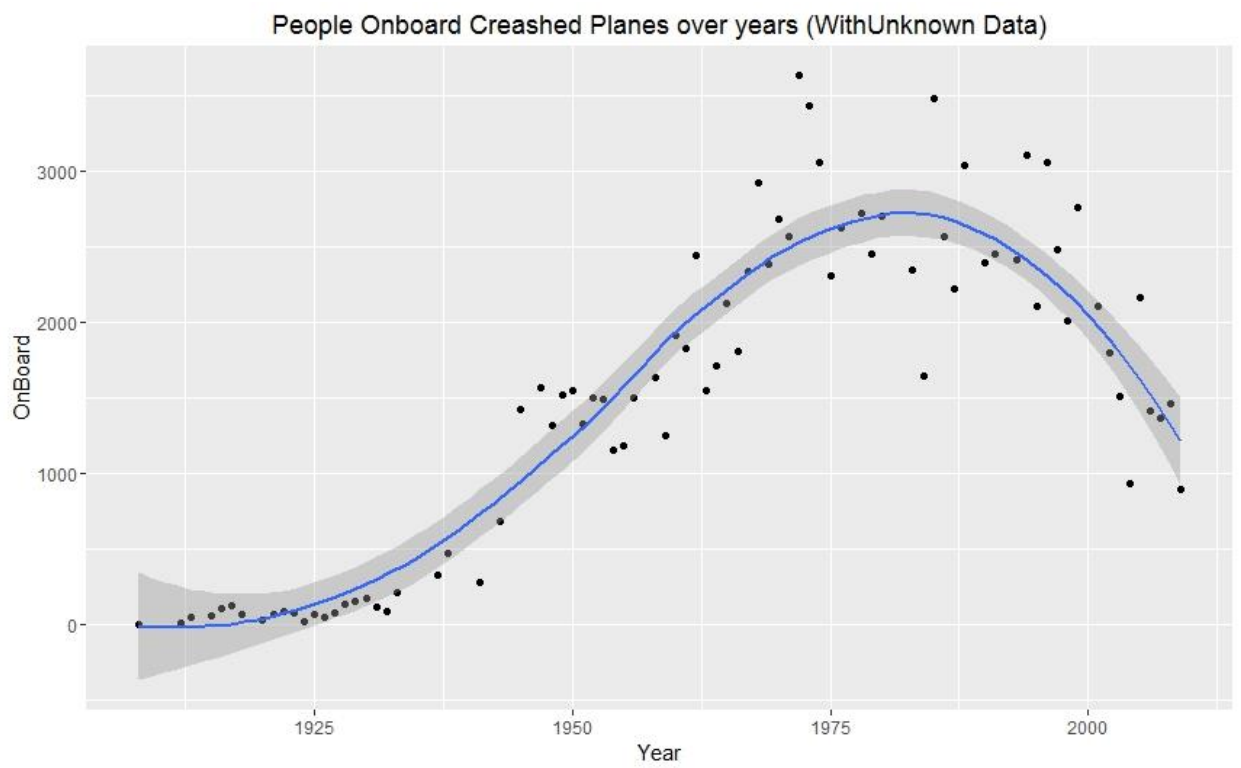
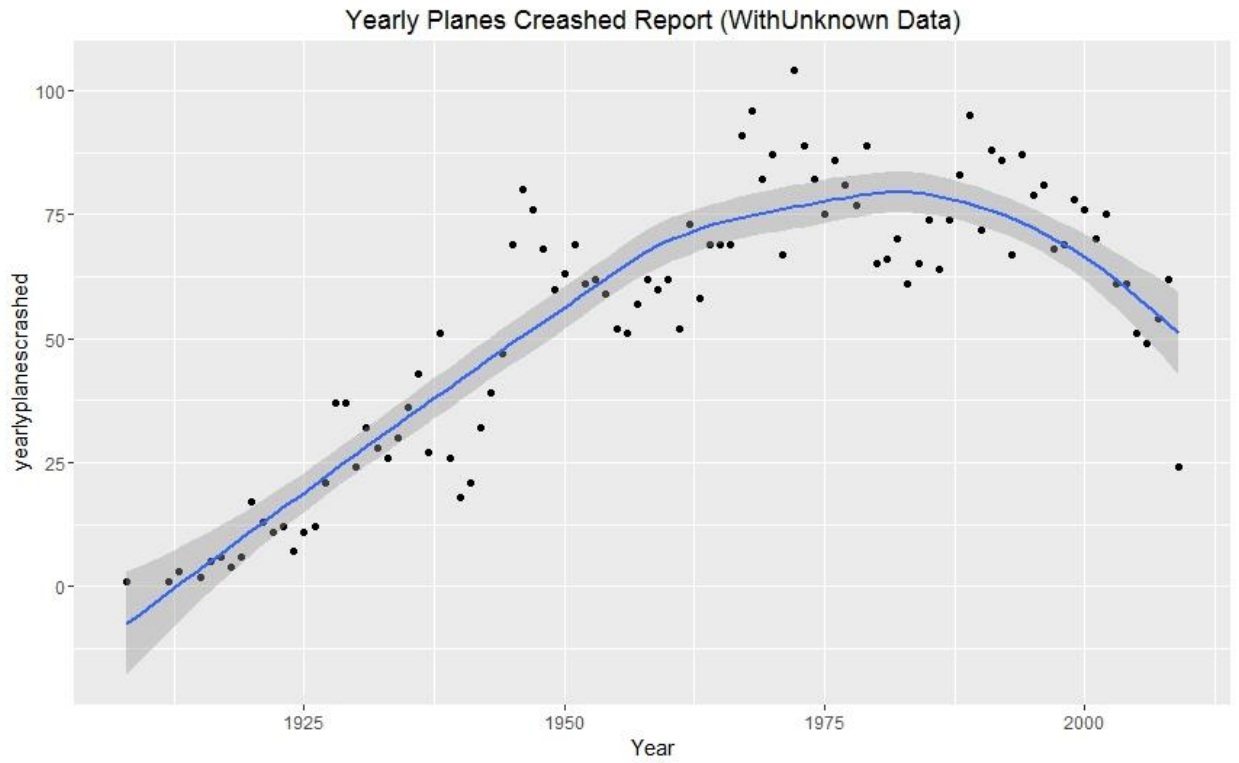


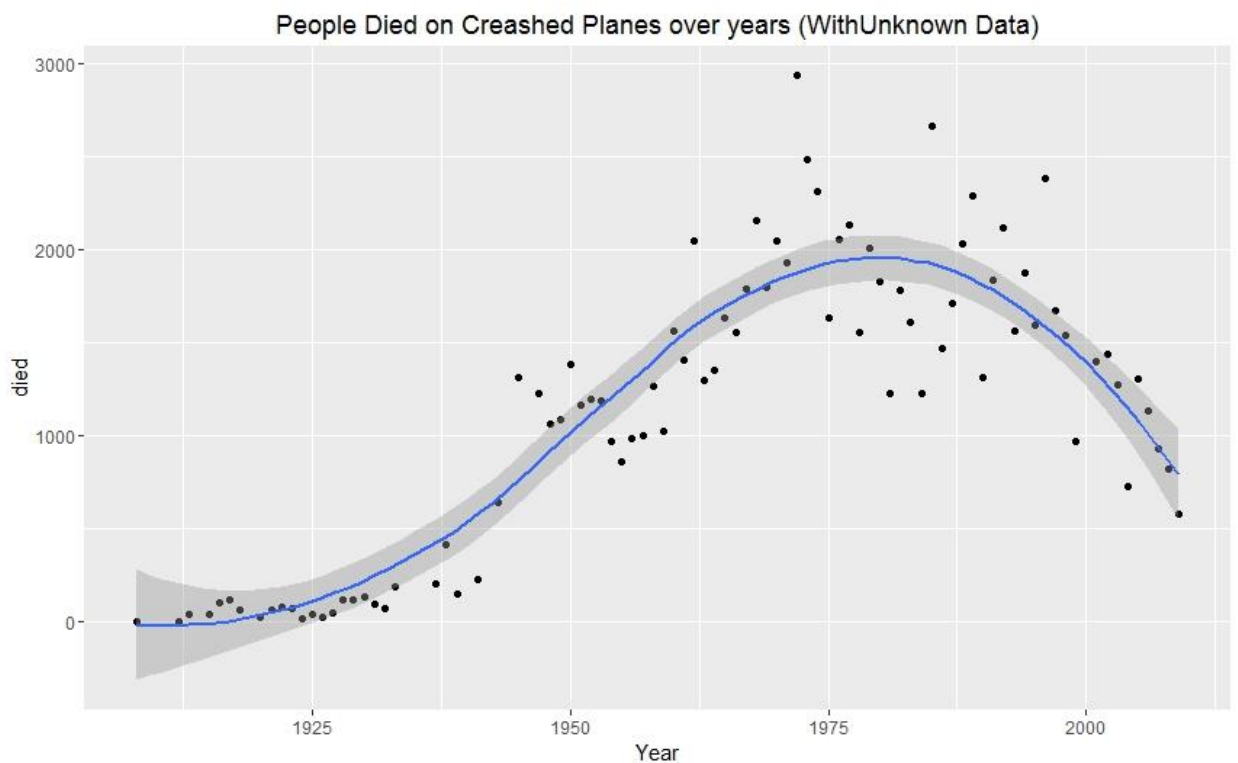
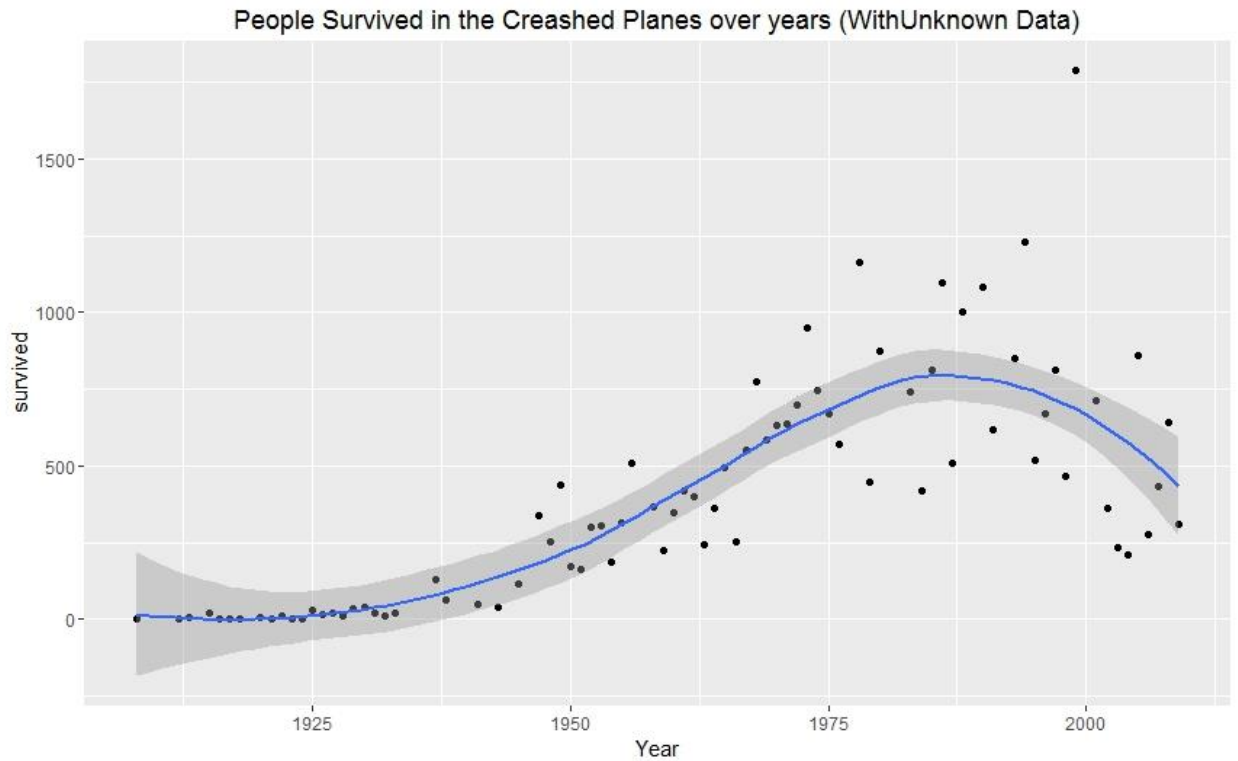
Here we can see that at the beginning the number of servicers were low then it started to increase till it reached the maximum level at the late 80's then started to retrieve again.



Number of recorded people died in plane crashes increased over time till it reached the maximum in the mid 70's which is same period for the most number of crashed planes. The maximum recorded number of death related to aircraft crashes was 2937 dead at the same year which has to most number of plane crashes 1972. We can see also that the number of death started to decrease as the number of crashed airplanes started to decrease. We can conclude also that there is a relation between number of airplane crashes and number of death.

- Result can be found in the "**YearlyPlanesCreashedReportWithUnKnown.csv**" file for the flights including some unknown data which are number of people onboard or number of people survived or died are missing. It's also proved that result with unknown values are almost the same with the known values, so results for unknown values will be neglected later.





As we can notice from the previous 4 pictures that almost there is no difference between result with and without the missing values or unknown data, so as mentioned before I'll neglect results for missing values, as it's a trivial percent of the data.

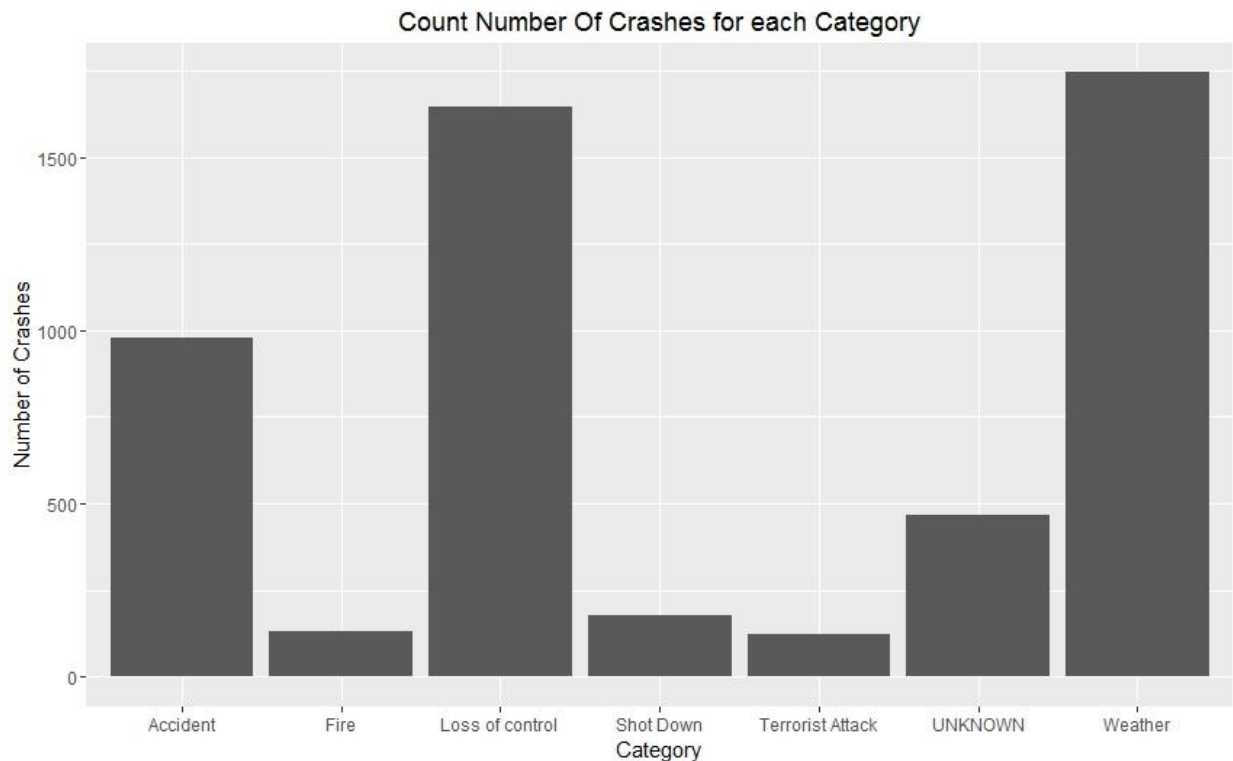
b- Highest number of crashes by operator and Type of aircrafts.

Result can be found in the **“CrashByOperatorandAircraftType.csv”** file for the flights crash report grouped by operator and type of aircraft.

c- ‘Summary’ field has the details about the crashes. Find the reasons of the crash and categorize them in different clusters i.e Fire, shot down, weather (for the ‘Blanks’ in the data category can be UNKNOWN) you are open to make clusters of your choice but they should not exceed 7.

The full result for the whole file summary field modified and replaced with the reason category can be found in **“PlaneCrashesAfterModifySummaryToCategories.csv”** file, and for the count numbers of crashes for each category can be found in **“CategoriesTableWithAccedentsCount.csv”** file and presented below.

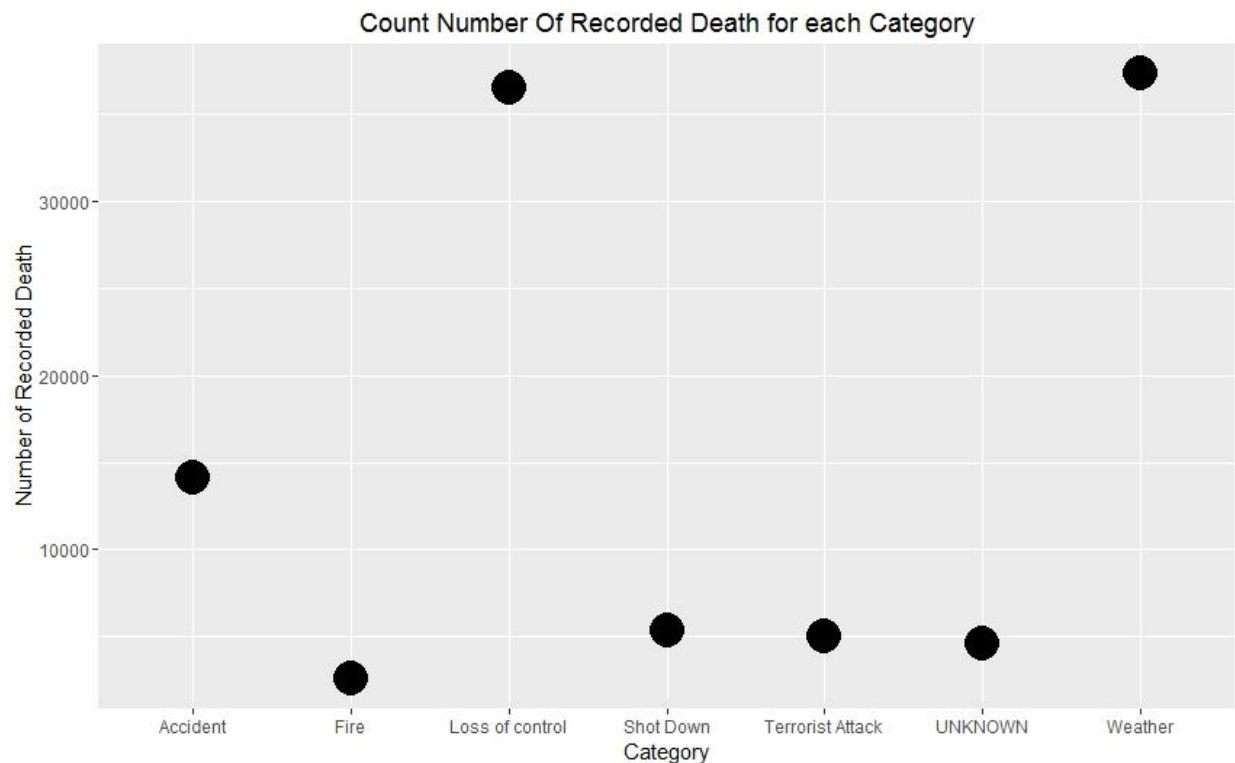
	<i>Category</i>	<i>Number of Crashes</i>
1	Weather	1746
2	Loss of control	1647
3	Accident	978
4	UNKNOWN	468
5	Shot Down	177
6	Fire	129
7	Terrorist Attack	123



- d- Find the number of crashed aircrafts and number of deaths against each category from above step.

Result for this point has been extracted to “**planeCrashesWithCategoriesandRcorderdDeath.csv**” file, and shown below.

	<i>Summary</i>	<i>Number of Crashes</i>	<i>Number of Recorded Deaths</i>
1	Weather	1746	37290
2	Loss of control	1647	36521
3	Accident	978	14090
4	Shot Down	177	5332
5	Terrorist Attack	123	5054
6	UNKNOWN	468	4582
7	Fire	129	2610



As appears from the previous presented data that the most reason for crashes is the weather conditions, and it has the highest number of death also. We can see that number of death is related to number of crashes whatever the reason is.

- e- Find any interesting trends/behaviors that you encounter when you analyze the data set.
- Number of crashes started to decrease from mid 90's.
 - Number of death is related to number of crashes.
 - Weather conditions are the most reason for aircraft crashes.
 - Loss of control almost has the same number of weather as a reason for crashes, and it has to analyzed and fixed.