5. Data Analysis

After completing data cleaning and data merging, the output file was useful for analysis using visualization tools. I used Tableau for this data analysis.

5.1 New Death Data with Imputed Records

The cleaned output file was analyzed with different cases such as with imputed data and without imputed data. The first data set I used for analysis was with imputed data records: New Death Data with 5610 records which included 767 imputed records. All analysis related to the lifetimes of satellites.

Lifetime1 = Death Date – Launch Date.

From Tableau with the New Death Data file, the following visualizations are observed.

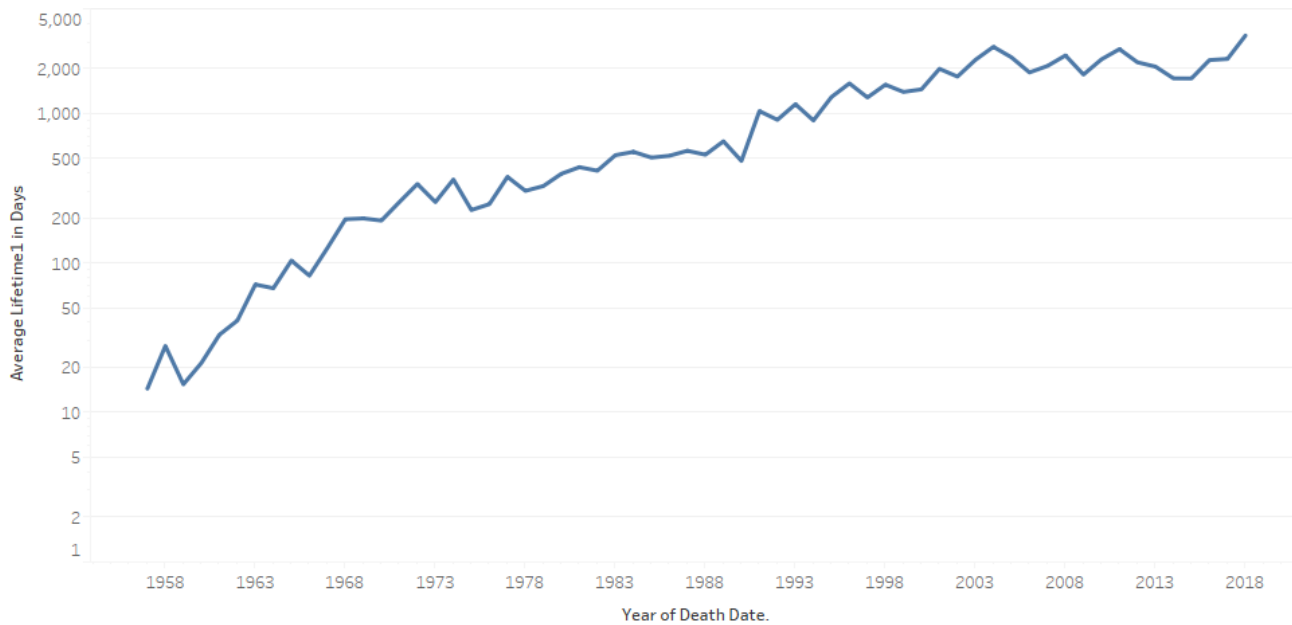

Figure 2: New Death Data: Average Lifetime in Days.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime in Days.

The above line chart was developed using Tableau. This chart shows the relation between average lifetime in days vs. year of death date. From the line chart, it was observed that from 1957 to 1990 the line increases gradually. The satellites that died between these years had small average lifespans of less than one year between 1958 to 1973 and average lifespans between one and two years from 1974 to 1990.

The lifetimes in months and years were also analyzed using Tableau and those visualizations are shown below in Figure 3a and Figure 3b. All these figures conceptually represent the lifespans of satellites.

Lifetime in months was based on year and month of satellite launch and death ignoring the days of the months. Lifetime in years were based on year of satellite launch and death, ignoring the month and day information

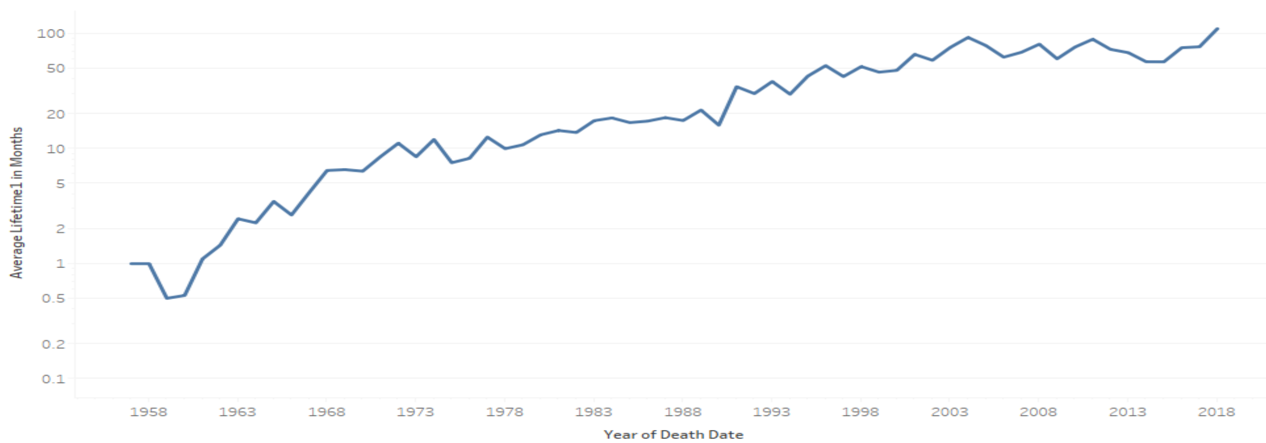

Figure 3a: New Death Data: Average Lifetime in Months.

X Axis Represents Year of Death date. Y Axis Represents Average Lifetime in Months.

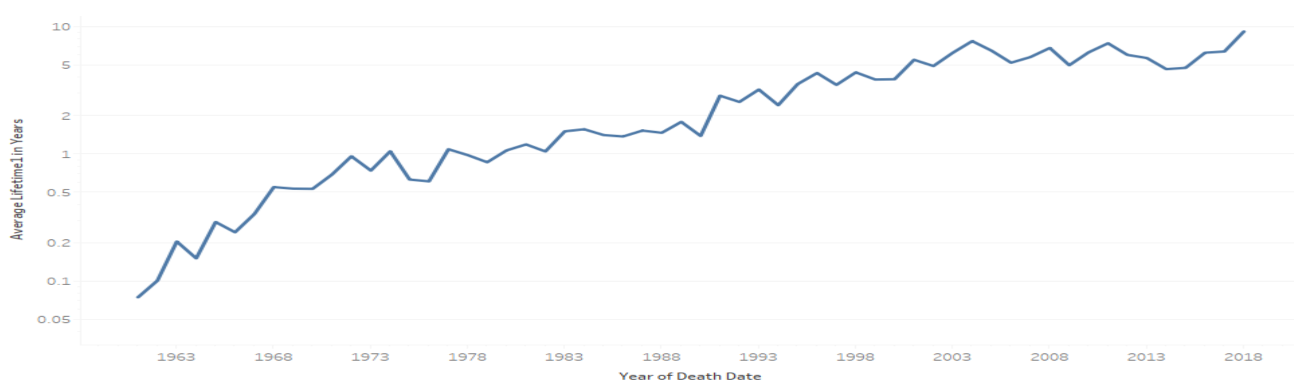

Figure 3b: New Death Data: Average Lifetime in Years.

X Axis Represents Year of Death Date. Y Axis Represents Average Lifetime in Years.

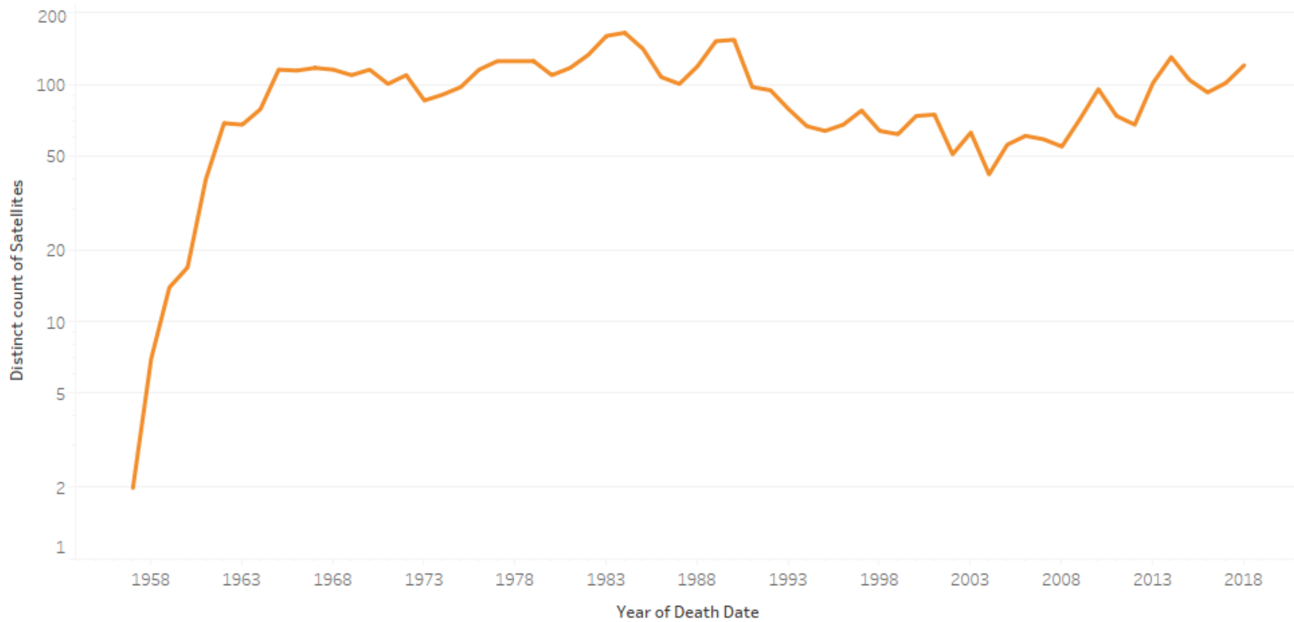

Figure 4: New Death Data: Distinct Count of Satellites.

X-Axis Represents Year of Death Date. Y-Axis Represents Distinct Count of Satellites (Total Satellites =5610)

Figure 4 represents distinct count of satellites vs. the year their lifespans ended. The count was increased rapidly from 1957 for several years and seems to level off thereafter, suggesting that satellite lifespans have been ending at a more or less constant rate for most of the space age so far.

The following below figures were developed with linear scale using Tableau.

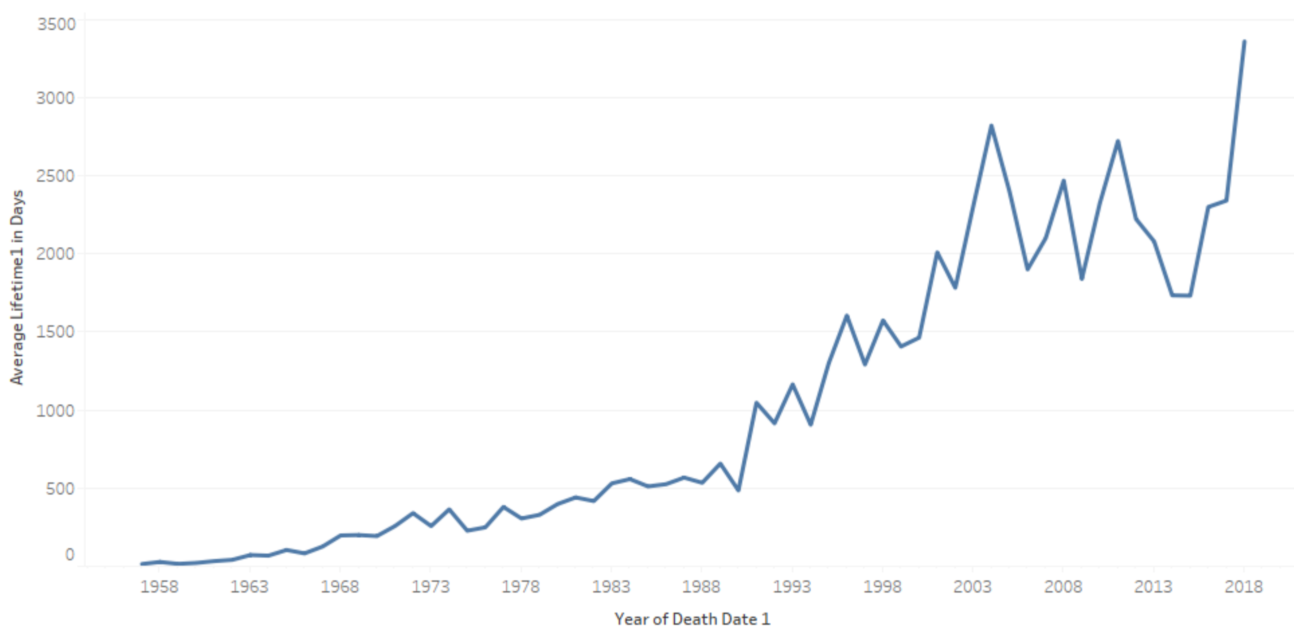

Figure 5: New Death Data: Average Lifetime in Days.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

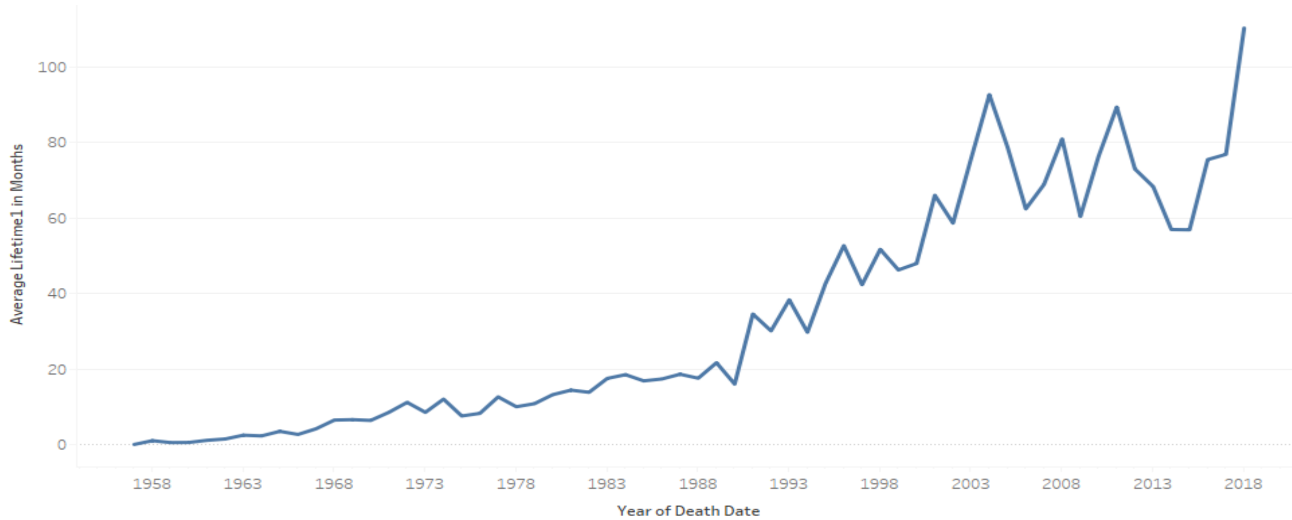

Figure 6: New Death Data: Average Lifetime in Months.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

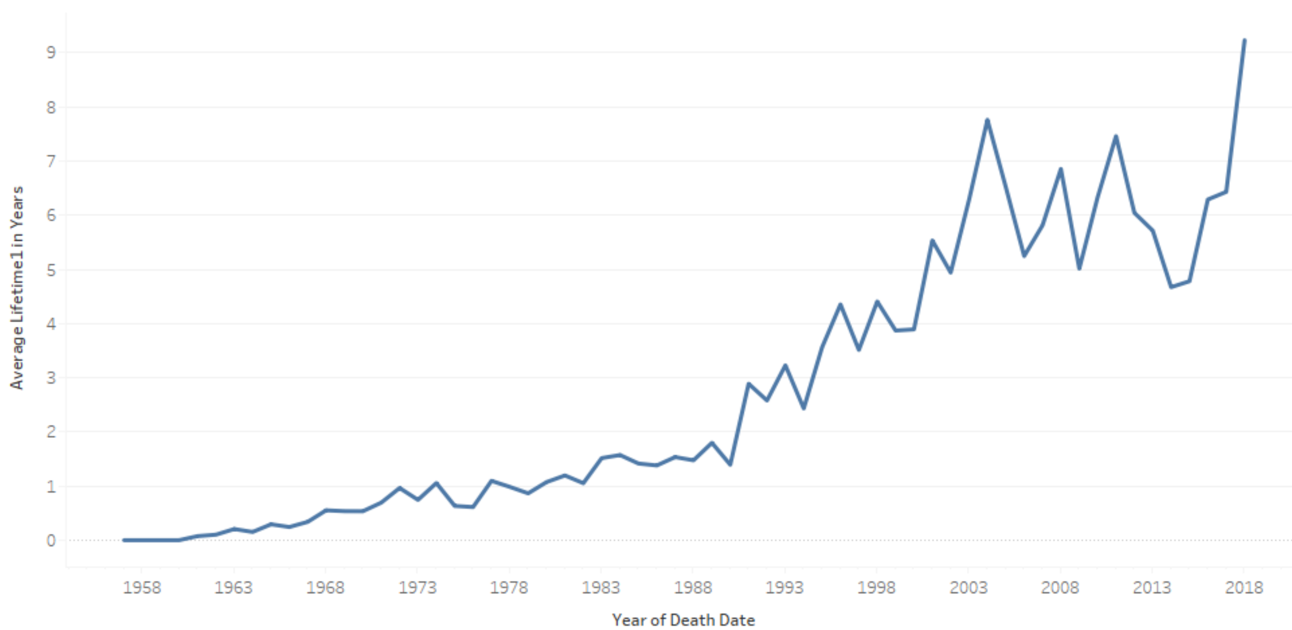

Figure 7: New Death Data: Average Lifetime in Years.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

From the linear graph analysis, it was observed that how gradually or rapidly the average lifetimes of satellites changed. Between 2002 and 2006 the lifetime escalates sharply and between 2011 to 2015, a steep descent was observed.

5.2 Imputed New Death Data Records

In the New Death Data some satellite data has Year and Month but is missing the day. Here are some estimation ideas to consider the missing day value.

If we have month but not the day of death,

- If launch was a different month, then estimated day of death is the middle of the month.
- If launch was in the same month, then estimated death is the middle of the rest of the month.
- If we have year but not the month or day, do not use these in the analysis.

Total New Death Data satellites with imputed death date = 767.

This data was used for the following graphs and developed using Tableau.

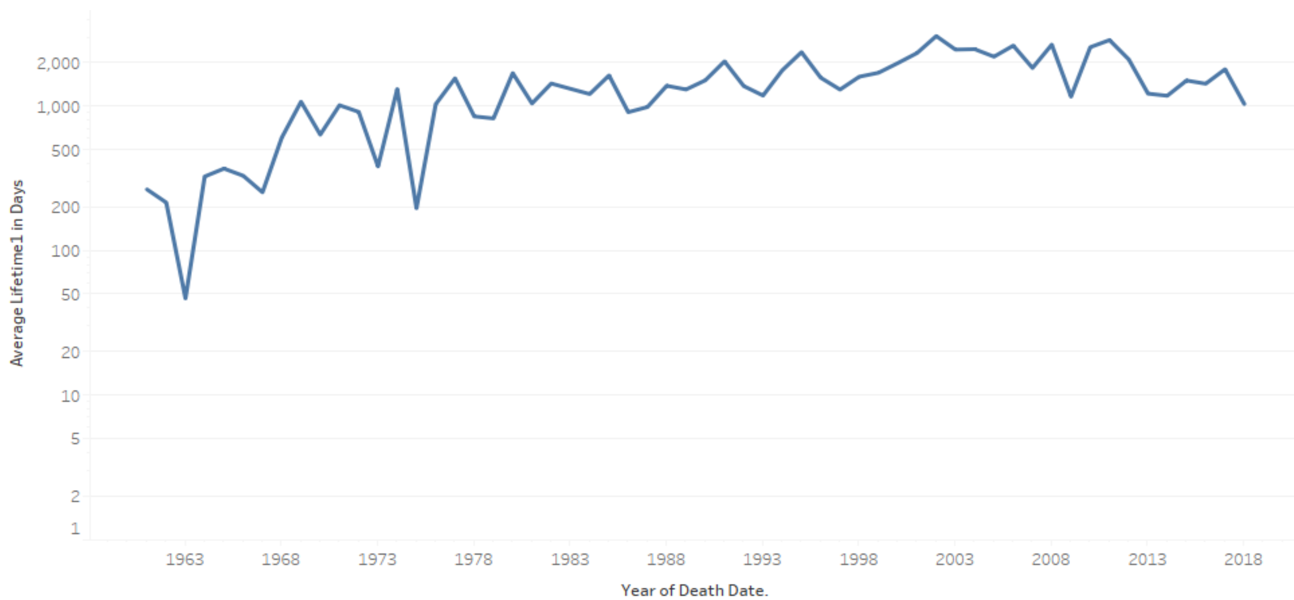

Figure 8: Imputed Data: Average Lifetime in Days.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

In this graph the average lifetime in days was measured with respect to death year of satellite. The average lifespan increases until 2002 and then it went down. In 1963 and 1975 the average lifetime was going down sharply.

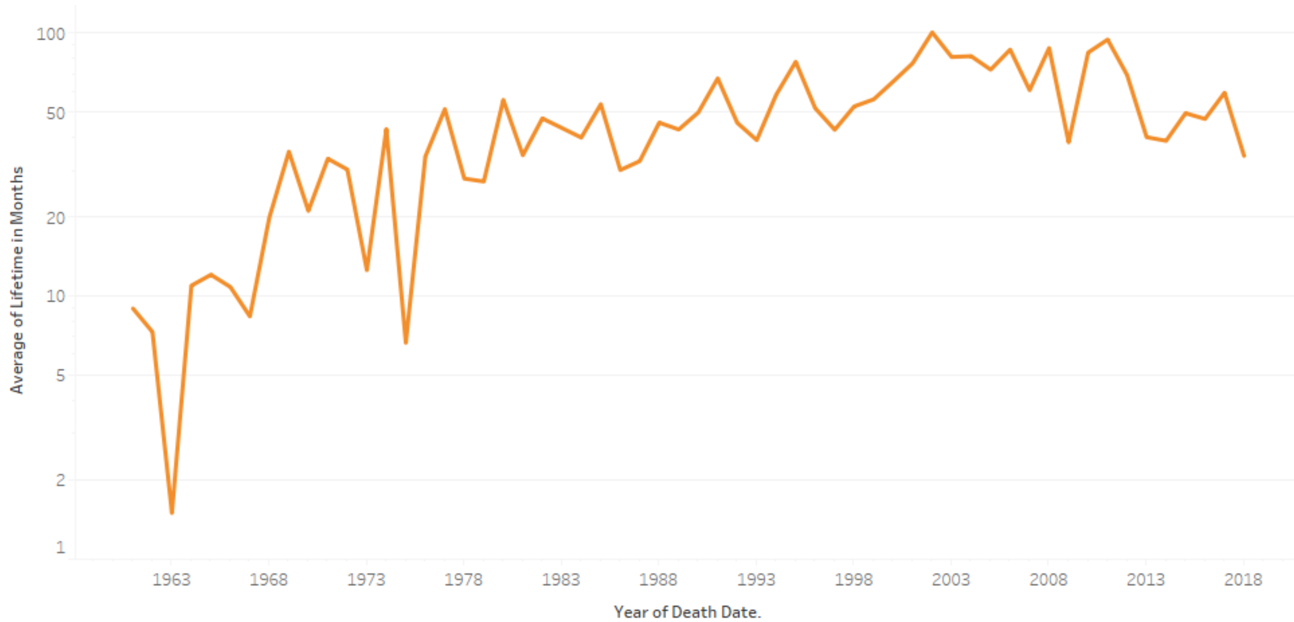

Figure 9: Imputed Data: Average Lifetime in Months.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

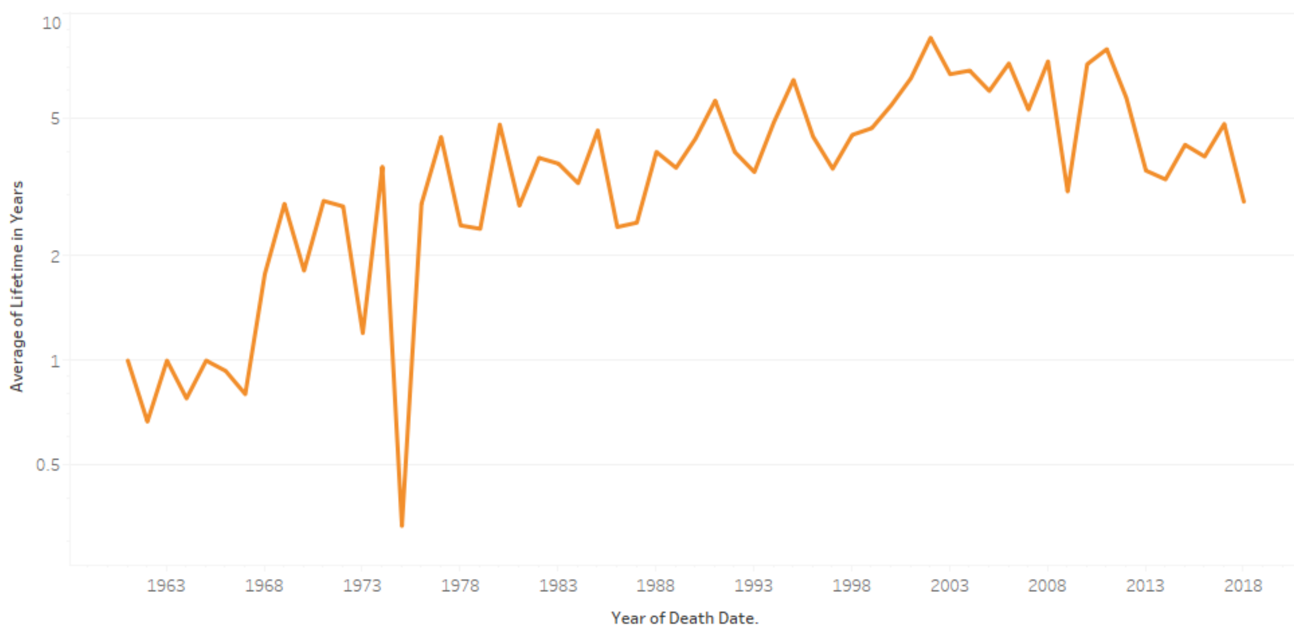

Figure 10: Imputed Data: Average Lifetime in Years.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

These graphs were measured in average lifetime in months and average lifetime in years with respect to death date. In Figures 8 to 10 the average lifetime increased exponentially until 2011 and from 2011, it was going down.

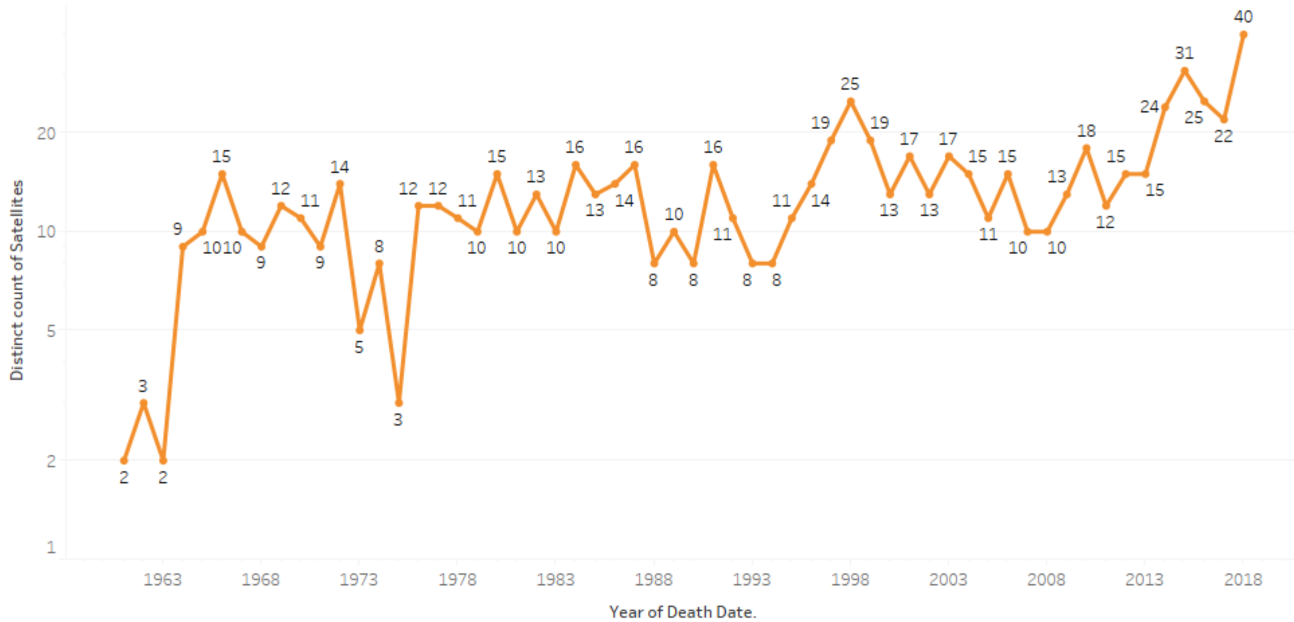

Figure 11: Imputed Data Satellite Count.

X Axis Represents Year of the Death Date. Y Axis Represents Distinct Count of Satellites with Imputed Death Dates.

Figure 11 shown the Distinct count of satellites with respect to Death Year and representing how many satellites were dead during that particular year.

5.3 New Death Data with All Values in Death Date (Year, Month and Day)

This data contains rows with all values in the Death Date field of New Death Data. Imputed death date satellites were eliminated from this analysis. Out of 9031 records I separated out 4843 records which have all values of year, month and day in Death Date field. This file was named New Death Data1.

Total satellites: 4843

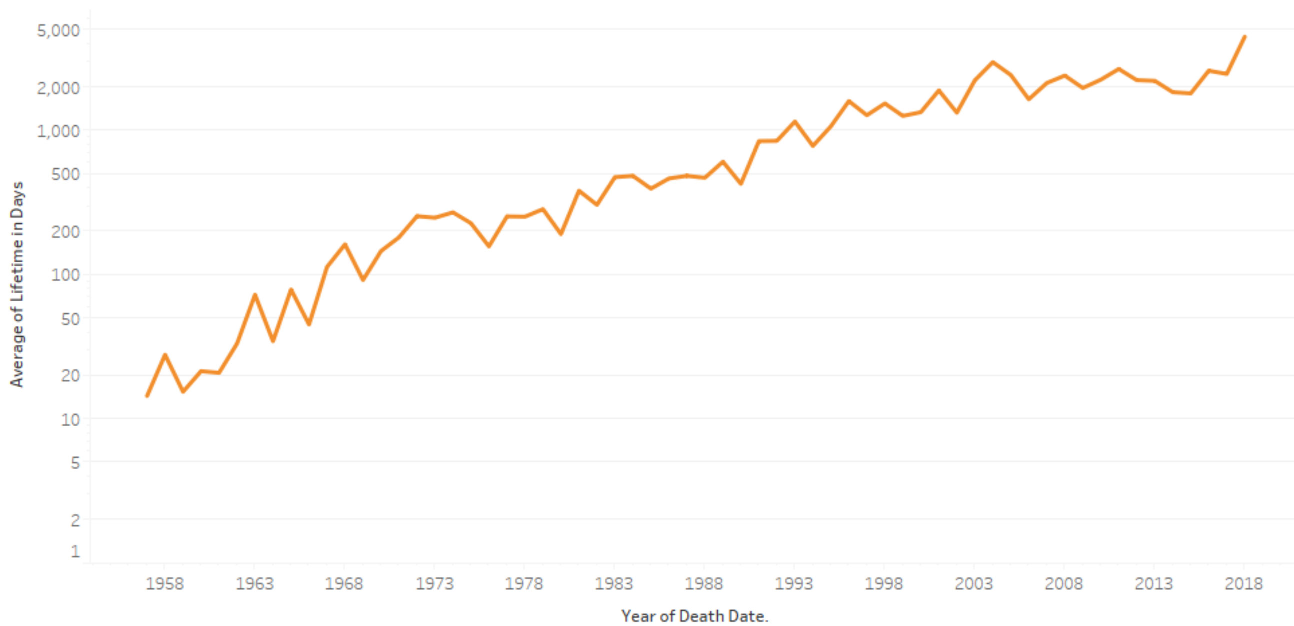

Figure 12: New Death Data1: Average Lifetime in Days.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

In Figure 12 the average lifetime in days was measured with respect to death year of satellite. The average lifespan increases until 2004 and then it varies slightly from 2014 to 2018.

Regarding Figures 12, 16, and 18, McDowell wrote (personal communication), “I did a similar calculation and came up with very similar graphs.”

In Figure 13 and Figure 14 the average lifetime was shown in days and years. From Figure 12 to Figure 14 has no difference in behavior of the graph when comparing with each other.

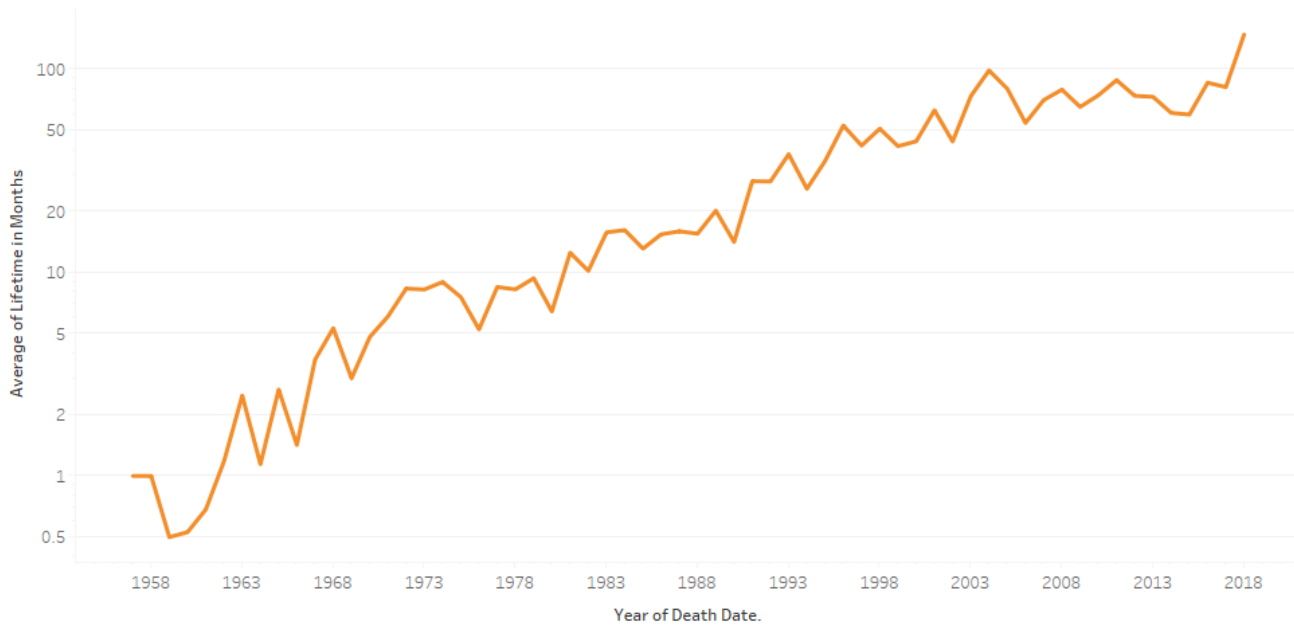

Figure 13: New Death Data1: Average Lifetime in Months.

X Axis Represents Year of the Death. Y Axis Represents Average Lifetime (Death Date – Launch Date).

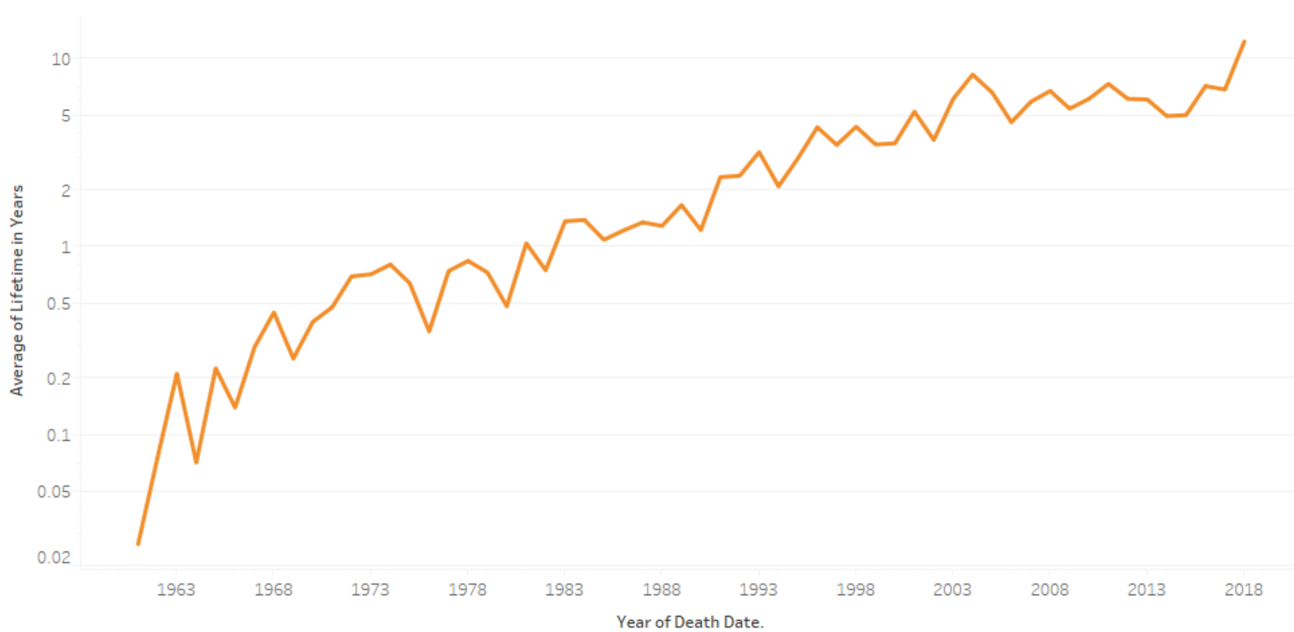

Figure 14: New Death Data1: Average Lifetime in Years.

X Axis Represents Year of the Death Date. Y Axis Represents Average Lifetime (Death Date – Launch Date).

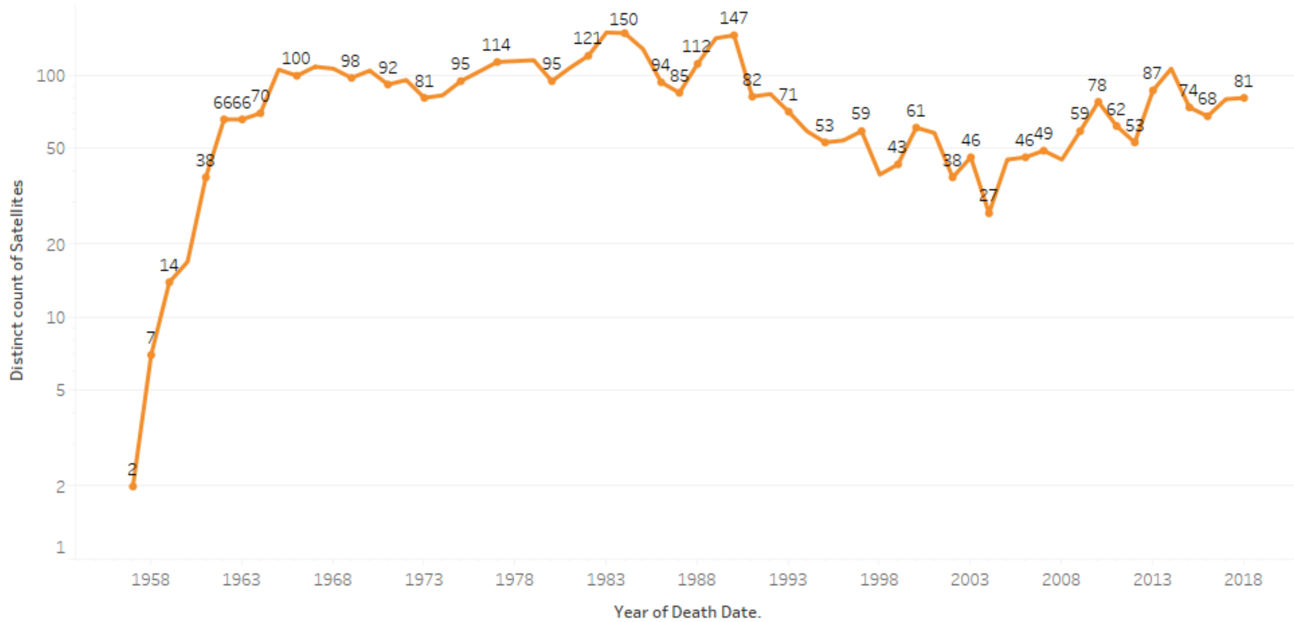

Figure 15: New Death Data1: Satellite Count.

X Axis Represents Year of the Death Date. Y Axis Represents Distinct Count of Satellites (Excluding Satellites with Imputed Days).

Figure 15 shown distinct count of satellites vs. New Death Data1. The count increases quickly until 1990 and thereafter the satellite count was going down.

5.4 New Death Data with Status Date as Reference

The status date column in SATCAT also can have missing values in the status date like year, month and/or day. I considered those satellites with status dates with all values (Year, Month and Day) in the date field.

In this analysis, I took Status Date as a proxy for death date.

JM's comment: However, more precisely, it is on average close to reentry date, if "In Earth Orbit" statuses are filtered out, as done here. According to McDowell (personal communication), "What you call "Date of Status" or "SATCAT status date" is better termed "decay date" or "reentry date". It's the date the satellite reentered or was deorbited.

The different statuses in the SATCAT data are shown in the below table.

Deep Space
Deep Space attach
Deorbited
Exploded
In Earth orbit
Landed
Landed Att
Lost
Reentered
Reentered Att
Spurious

Table 10: SATCAT Current Statuses

From New Death Data satellites with Status date given in SATCAT, the current status column was filtered to remove satellites with In Earth Orbit status. The filtered data was analyzed using Tableau.

Total satellites = 3586.

Two lifetime estimates were presented in this analysis.

- Lifetime1 = Death Date – Launch Date.

  (In the below figures **Average Lifetime1 in Days**, Months, Years represents this calculation.)

- Lifetime2 = Status Date – Launch Date.

(In the below figures **Average Lifetime2 in Days**, Months, Years represents this calculation.)

The analysis is based on the following conditions.

- Average of Lifetime1 / Average of Lifetime2 vs. Death Year / Launch Year / Status Year
- Satellite count vs. Death Year / Status Year

From Figure 16 to Figure 30, the orange color line charts **Lifetime1 = Death Date – Launch Date** and the blue color line charts **Lifetime2 = Status Date – Launch Date.**

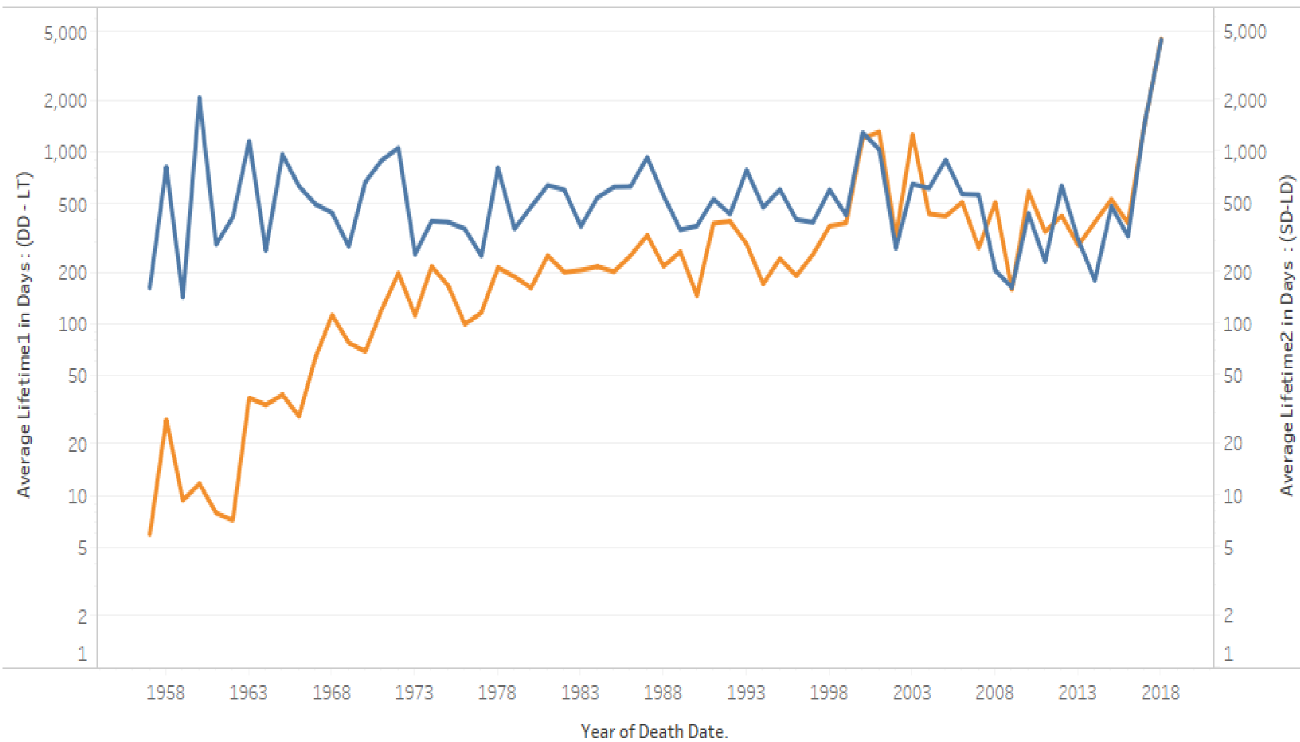

Figure 16: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime in Days Vs. Death Year. X Axis Represents Year of the Death Date. Y Axis Represents Yearly Average Lifetime1 and Yearly Average Lifetime2.

Figure 16 shows average lifetime in Days with respect to Death Year. The Average Lifetime1 increases gradually from 1958 to 1999. From 2000 to 2016 it went down and again it went up from 2017. Lifetime2 with respect to Year of Death was not like Lifetime1.

From 2000, Lifetime1 and Lifetime2 were relatively close and the Average Lifetime went down from 2000. From 2000, the difference between average of lifetime1 and lifetime2 was smaller.

From this chart, Lifetime1 with Death Date has some meaningful observations. The highest Lifetime1 was 4588 days and Lifetime2 was 4478 days.

Regarding Figures 12, 16, and 18, McDowell wrote (personal communication), “I did a similar calculation and came up with very similar graphs.” Also, “The lower values for LT1 vs Y1 in Fig 16 vs Fig 12 are not surprising. You have removed all satellites which are dead but are still in orbit (LT1 is defined but LT2 is not defined). For various reasons these are likely to be longer lived satellites than those which are dead and down (which e.g. includes a lot of short lived recoverable missions like human spaceflight missions and early spy sats).” He commented, “You could partly verify this by making a version of Fig 12 using only the sats *not* included in Fig16.”

According to McDowell, “in Fig 16, LT2 (average time to reentry) is independent of death date and is of order a year. This is saying that the majority of satellites in LEO reenter naturally rather than being actively deorbited. Therefore their orbital lifetime depends mostly on the atmosphere, not on the technology of the satellite or whether it is working or not. LT1 increases from 1957 to about 1995. This does reflect an increase in satellite reliability: in the 1960s most LEO satellites failed quickly, in the 2000s most LEO satellites worked until they burnt up on reentry (so LT1=LT2)”

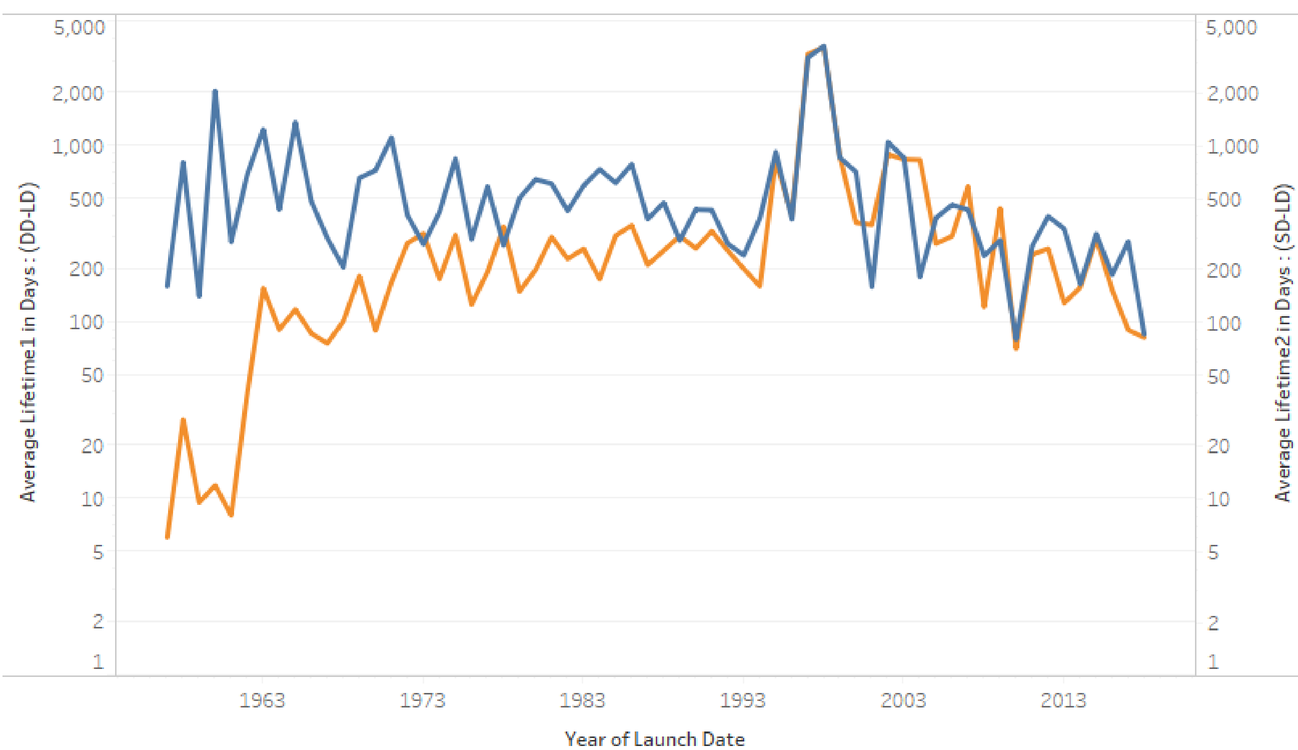

Figure 17: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime in Days Vs. Launch Year. X Axis Represents Year of Launch Date. Y Axis Represents Average of Lifetimes of Satellites Ending in a Given Year. Orange: Lifetime1, Blue: Lifetime2.

In Figure 17, the average lifetime is measured with respect to Launch Year of satellite. The average Lifetime1 increases until 1998 and then it decreases. But the average Lifetime2 with respect to Launch Year has fluctuations in the graph and from my thoughts, Launch Year was not suitable for analyzing the Lifetimes of satellites.

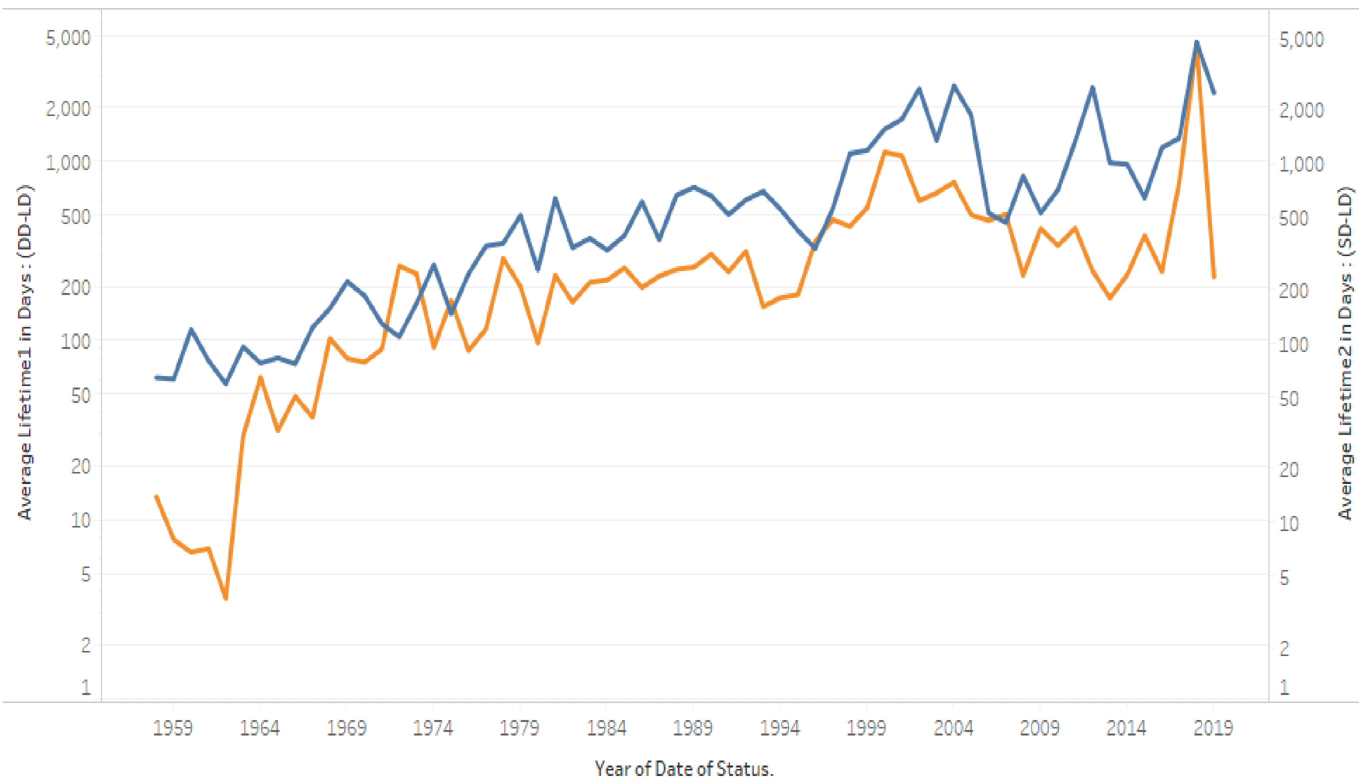

Figure 18: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime in Days Vs. Status Year. X Axis Represents Year of Status Date. Y Axis Represents Average of lifetimes of Satellites Ending in a Given Year. Orange: Lifetime1, Blue: Lifetime2.

Figure 18 shows the average lifetime measured with respect to Status Year of the satellite. Lifetime1 increases gradually until 2000 and then decreases. Lifetime2 also increases gradually and both lifetime graphs were similar but differ in average lifetime value. From this analysis, it was observed that Lifetime2 values were greater than Lifetime1 values, but both had the same pattern. Regarding Figures 12, 16, and 18, McDowell wrote (personal communication), “I did a similar calculation and came up with very similar graphs.” He also commented, “In Fig 18, you note LT2 is greater than LT1. Of course it is. LT2 represents reentry. The satellite can't keep working after reentry, so for each individual satellite LT1 <= LT2., hence this is also true for the average.”

According to McDowell, “You note that the blue curve has a different shape in Fig 18 vs Fig 16. In particular, we see that in Fig 18 the orbital lifetime to reentry correlates with reentry date. Why is this? It's because the total number of satellites in orbit increases with time. Very few sats which decayed in 1965 were launched before 1965, because there just weren't very many satellites at all before 1965. In contrast, in 1985, there were a lot of satellites from the 1970s reentering, as well as the short lived ones launched that year. But in Fig 16, the entry for 1965 includes satellites with long orbital lifetimes - ones that reentered in the 1970s, they just failed soon after launch. To say it another way, a satellite which is launched in 1965 and then immediately fails (short LT1), but stays in orbit for many years afterwards (long LT2) contributes to Fig 16 on the left (x axis 1965) but to Fig 18 on the right (x axis 1970s). Long orbiting sats squish Fig 18 rightwards.

From Figures 16, 17 and 18, Lifetime1 trended up and Death Year as the x-Axis and same with Launch Year and Status Year. But for Lifetime2, Death Year and Launch year for the x-Axis did not trend up and, but for Status Year, it did.

There were sudden variations in lifetime between 1996 to 2006 and from 2006 to 2016. Figure 19 to Figure 22 show average lifetimes each year for 10 years. Here it was from 1996 to 2006 and from 2007 to 2016.

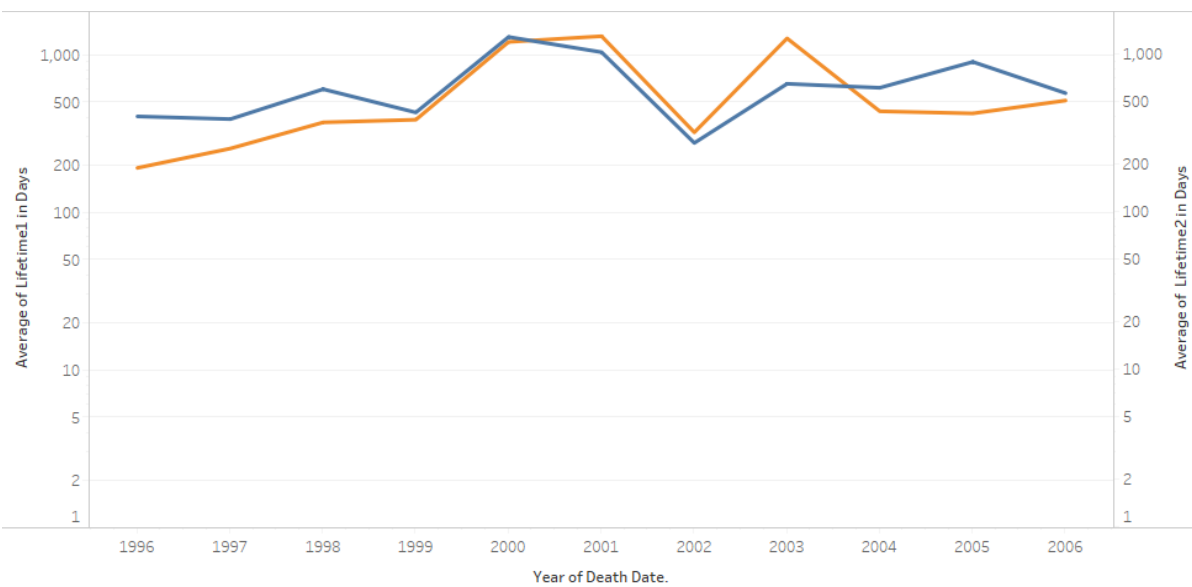

Figure 19: Yearly Average Lifetimes in Days (1996-2006) Vs. Year of Death Date.

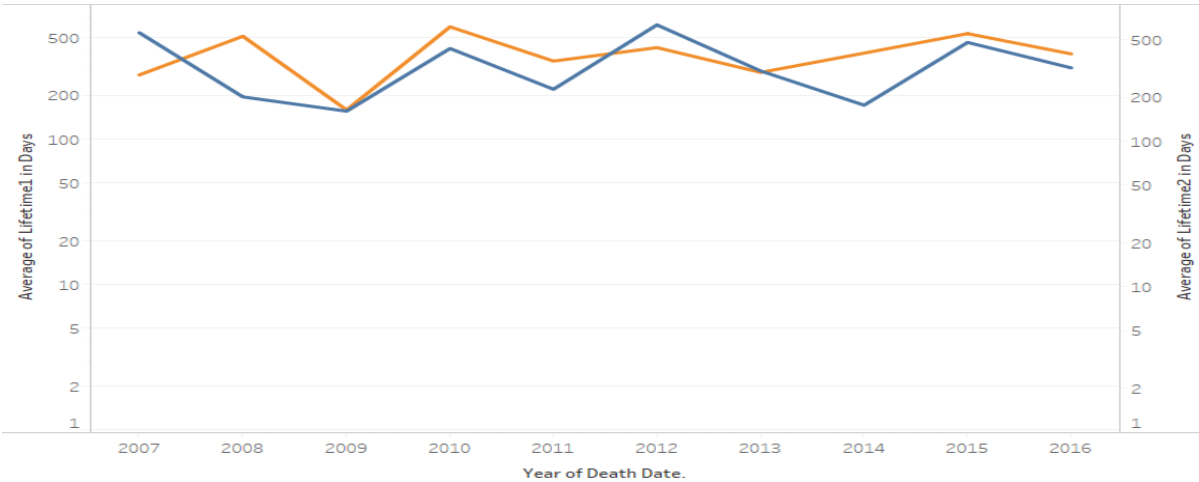

Figure 20: Yearly Average Lifetimes in Days (2007-2016) Vs. Year of Death Date.

The above two figures show an average for Lifetime1 and Lifetime2 in Days with respect to Death Year. From this analysis, it was observed that the average lifetime of satellites between 1996 to 2006 is higher when compared with average lifetime of satellites between 2007 to 2016. Both line charts have the same pattern with not much difference between them.

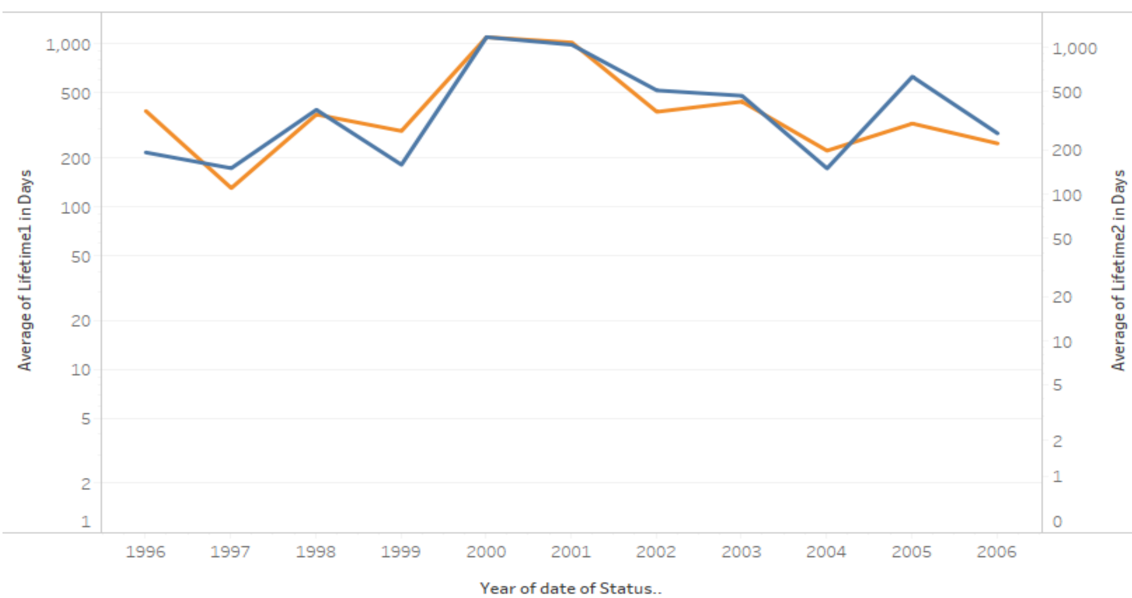

Figure 21: Average Lifetime in Days (1996-2006) Vs. Year of Status Date.

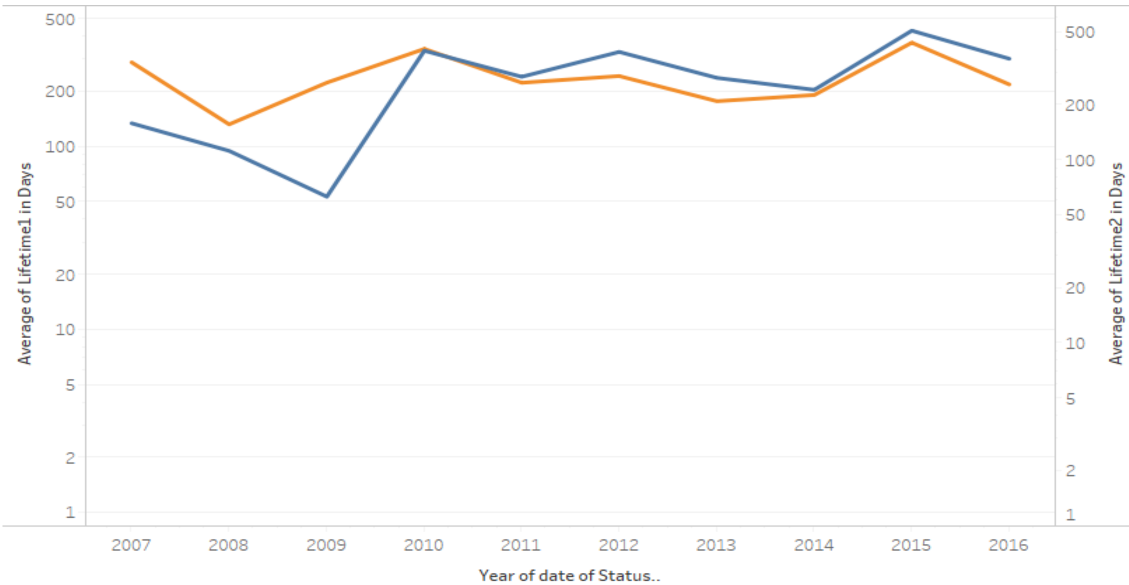

Figure 22: Average Lifetime in Days (2007-2016) Vs. Year of Status Date.

Figure 21 and 22 show the average of both lifetimes during the periods of 1996 to 2006 and 2007 to 2016 with respect to Status Year. Here also both lifetimes followed a similar pattern and are relatively close to each other, but I am still not sure that Status Date is a good proxy or not for the Death Date.

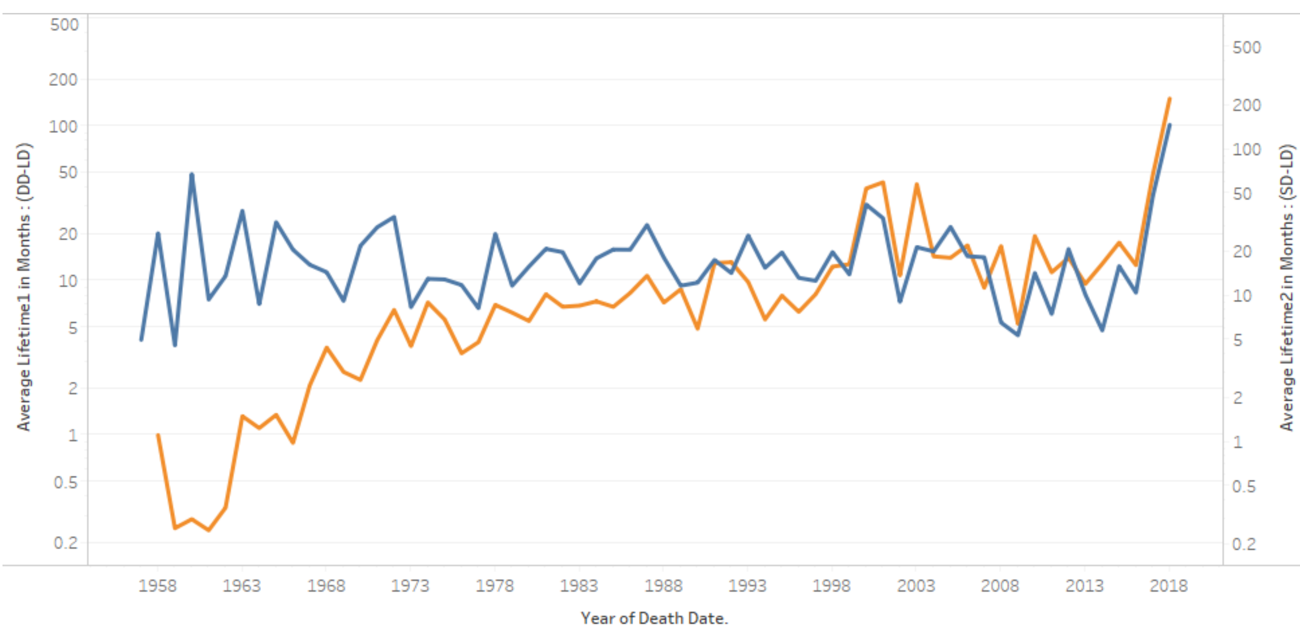

Figure 23: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime in Months Vs. Death Year. X Axis Represents Year of Death Date. Y Axis Represents Yearly Averages of Lifetime1 and Lifetime2. Orange: Lifetime1, Blue: Lifetime2.

Figure 23 shows Lifetime1 increasing until 1999 and thereafter it decreased. Lifetime2 with death date was harder to understand. But the following figures show some relation between the periods 1996 to 2006 and 2007 to 2016.

Figure 24: Average Lifetime in Months (1996-2006 & 2007-2016) Vs. Year of Death Date.

From this analysis, Lifetime2 with respect to Death Year is a curve that is hard to interpret.

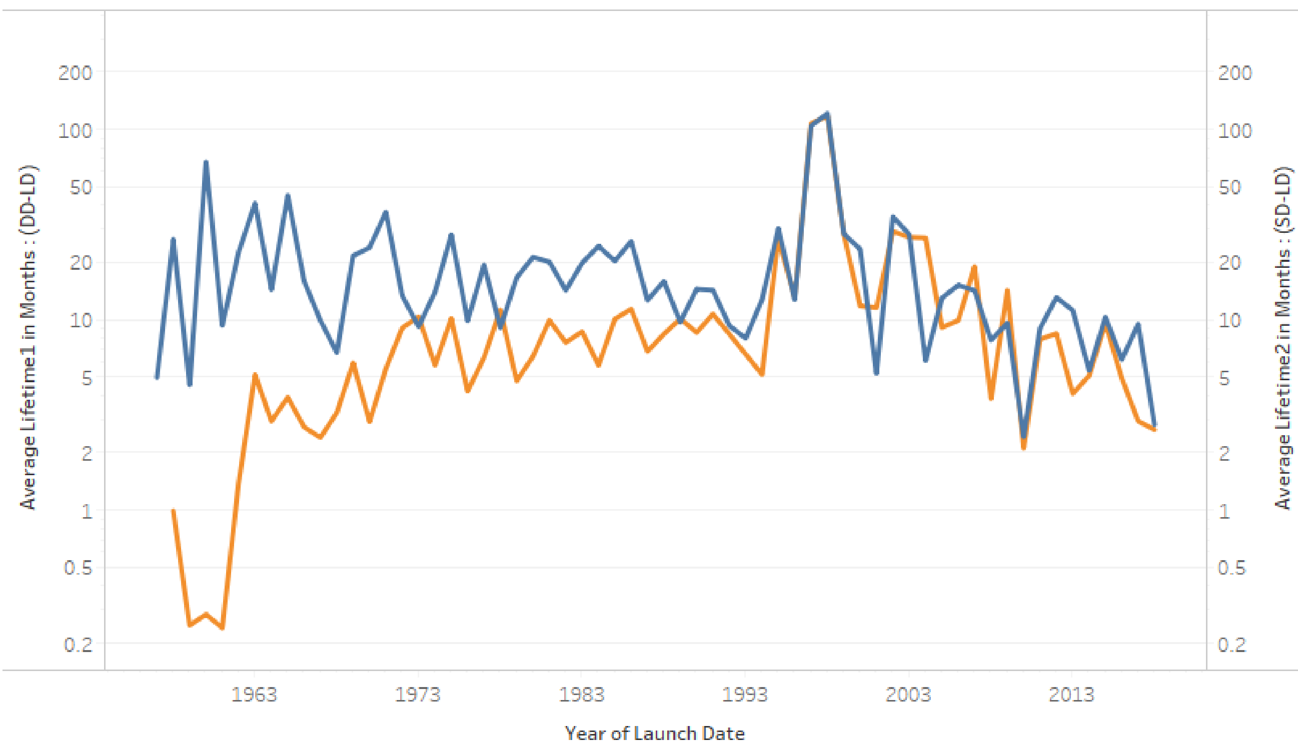

Figure 25: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime in Months Vs. Launch Year. X Axis Represents Year of Launch Date. Y Axis Represents Average of Lifetime1 and Lifetime2.

In Figure 25 the average lifetime in months is shown with respect to Launch Year of satellite. It is similar but not identical to Figure 17 which showed lifetimes in days. The average of Lifetime1 increases until 1998 and then it decreased. But the Lifetime2 with respect to Launch Year has little pattern because of fluctuations in the graph. My thought was that Launch Year was not suitable for the x-Axis to analyze lifetime data.

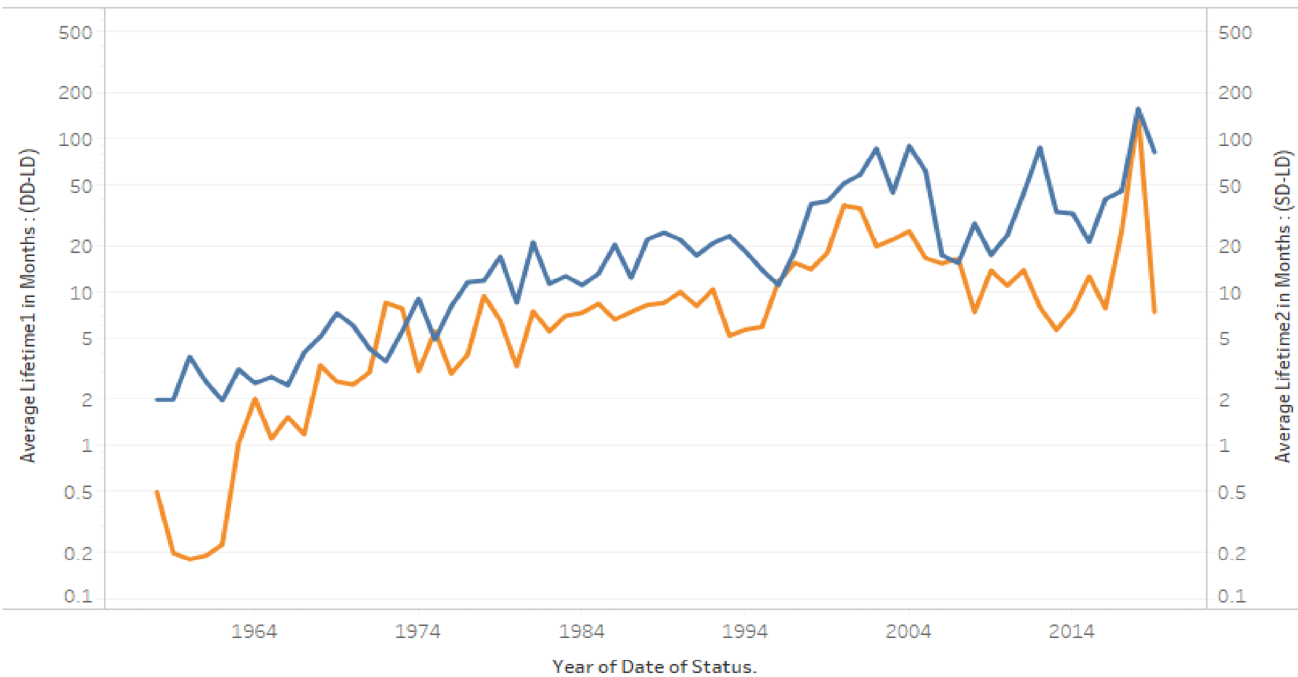

Figure 26: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime2 in Months Vs. Status Year. X Axis Represents Year of Status Date. Y Axis Represents Average Lifetime.

Figure 26 shows Lifetime1 and Lifetime2 of satellites with respect to Status Year in months (compare to Figure 18 which uses days). Both the lifetimes were close to each other and increased gradually up to year 2000. The analysis shows Lifetime1 and Lifetime2 vs. Status Year having relative closeness but Lifetime2 was greater than Lifetime1.

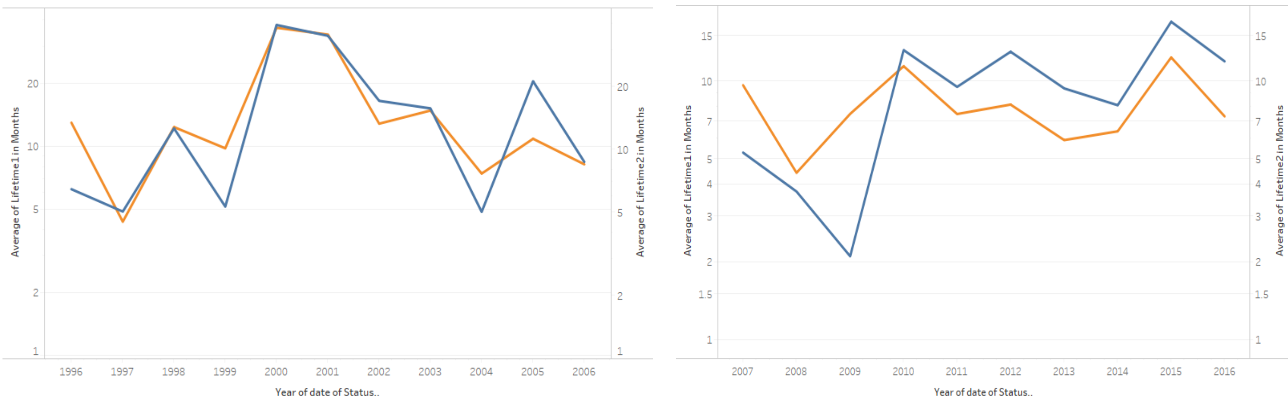

Figure 27: Average Lifetime in Months (1996-2006 & 2007-2016) Vs. Year of Status Date.

Figure 27 shows the average lifetime in months with respect to Status Year. In both cases in Figure 27 the lifetimes are following similar paths.

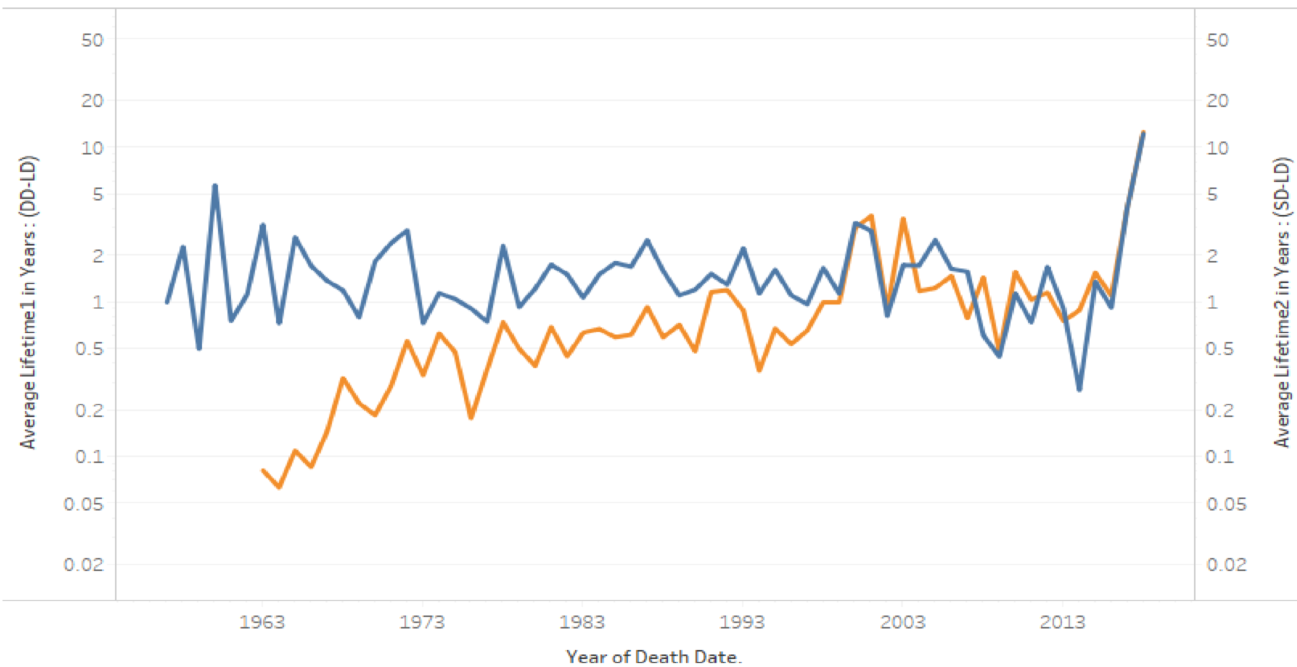

Figure 28: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime in Years Vs. Death Year. X Axis Represents Year of Death Date. Y Axis Represents yearly average for Lifetime1 and Lifetime2.

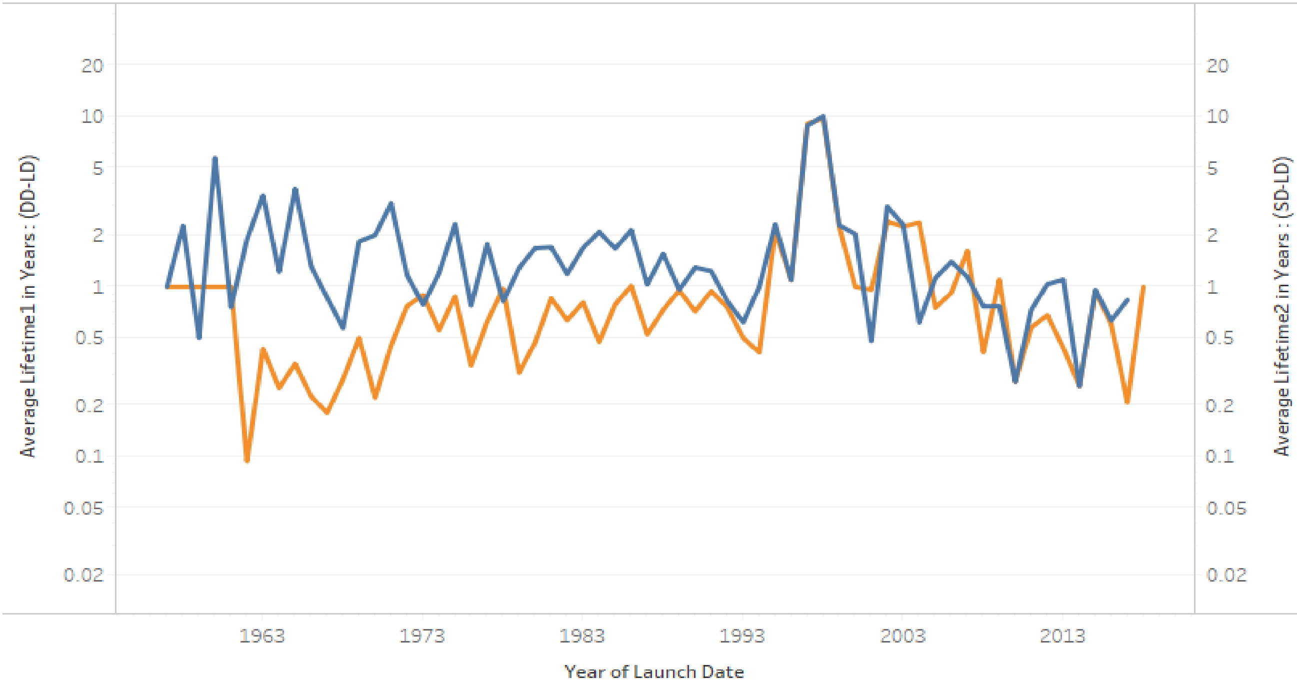

Figure 29: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime in Years Vs. Launch Year. X Axis Represents Year of Launch Date. Y Axis Represents Yearly Average for Lifetime1 and Lifetime2. The early part of the orange curve shows as 1 but is really 0. Orange: Lifetime1, Blue: Lifetime2.

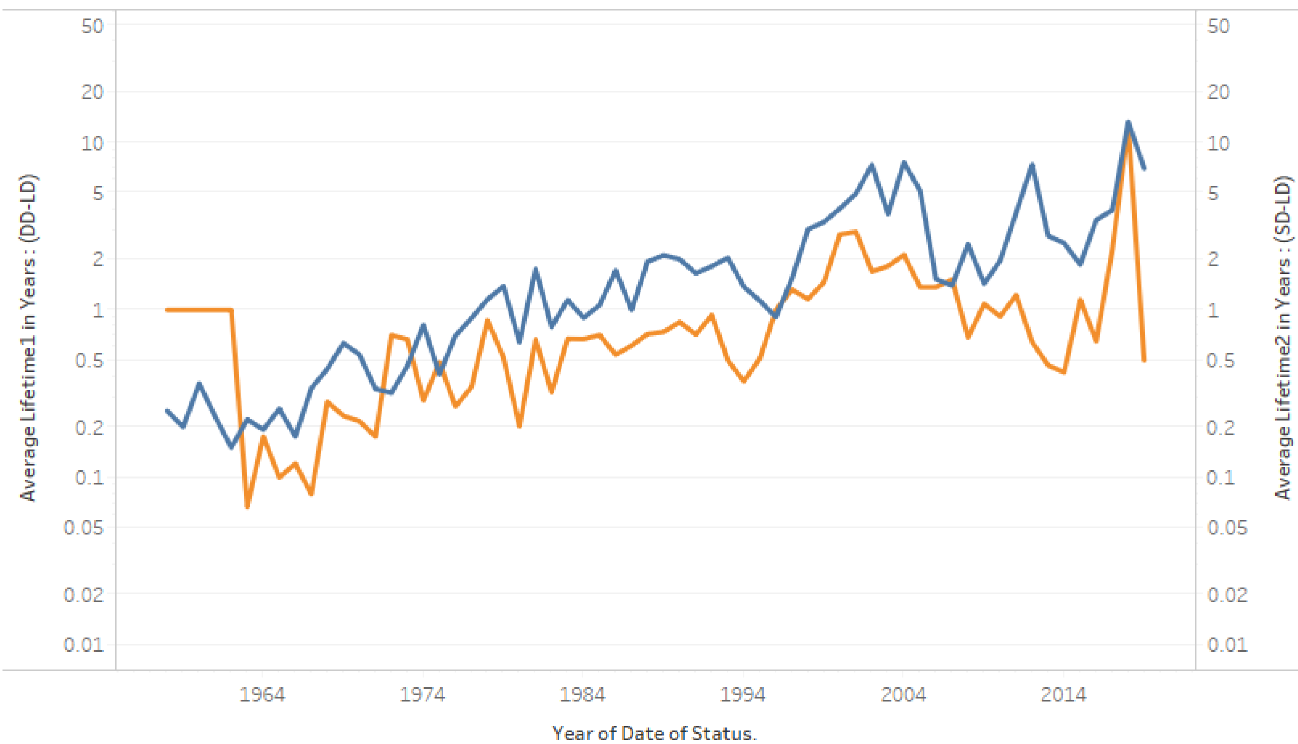

Figure 30: New Death Data Satellites with Status dates in SATCAT: Average Lifetime in Years Vs. Status Year. X Axis Represents Year of Status Date. Y Axis Represents Yearly Average for Lifetime1 and Lifetime2. The early part of the orange curve shows as 1 but is really 0. Orange: Lifetime1, Blue: Lifetime2.

Figure 28, 29 and 30 show average lifetimes in years with respect to Death Year, Launch Year and Status Year respectively. In all cases Lifetime1 seems to have exponential growth until the year 2000 and thereafter it changes. Both lifetimes with Status Year seem to have exponential growth until 2000.

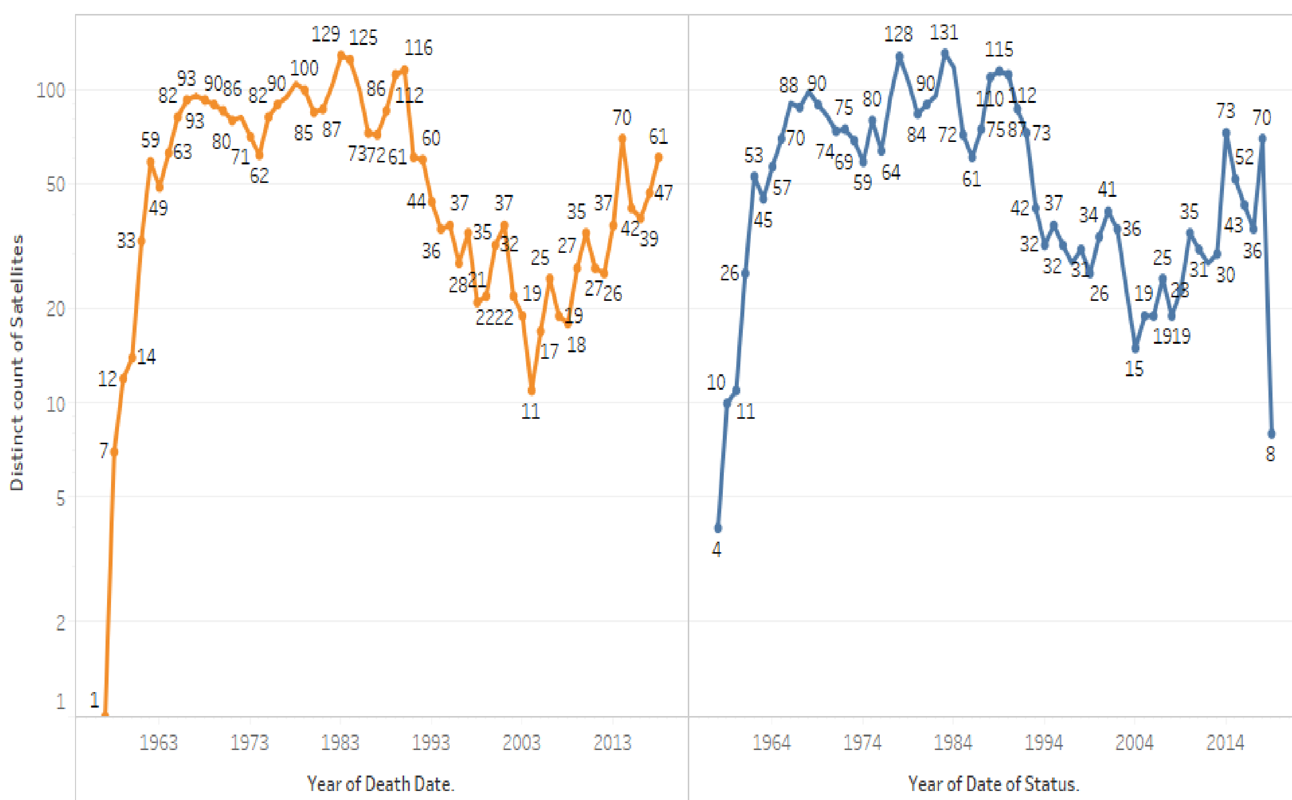

Figure 31: New Death Data Satellites with Applicable Status Dates in SATCAT: Distinct Count of Satellites Vs. Death / Status Year. X Axes Represent Year of Death and Year of Status Date. Y Axis Represents Distinct Counts of Satellites. Orange: Distinct Counts of Satellites vs. Death Year, Blue: Distinct Counts of Satellites vs. Status Year.

Figure 31 shows the satellite count with respect to Death Date and Status Date. In both cases the distinct count increased quickly and then decreased.

The figures above that use New Death Data satellites with Status date in SATCAT other than In Earth Orbit status show comparisons of how lifetimes with respect to Status Date, Launch Date and Death Date behave and how much similarity exists between the two lifetimes.

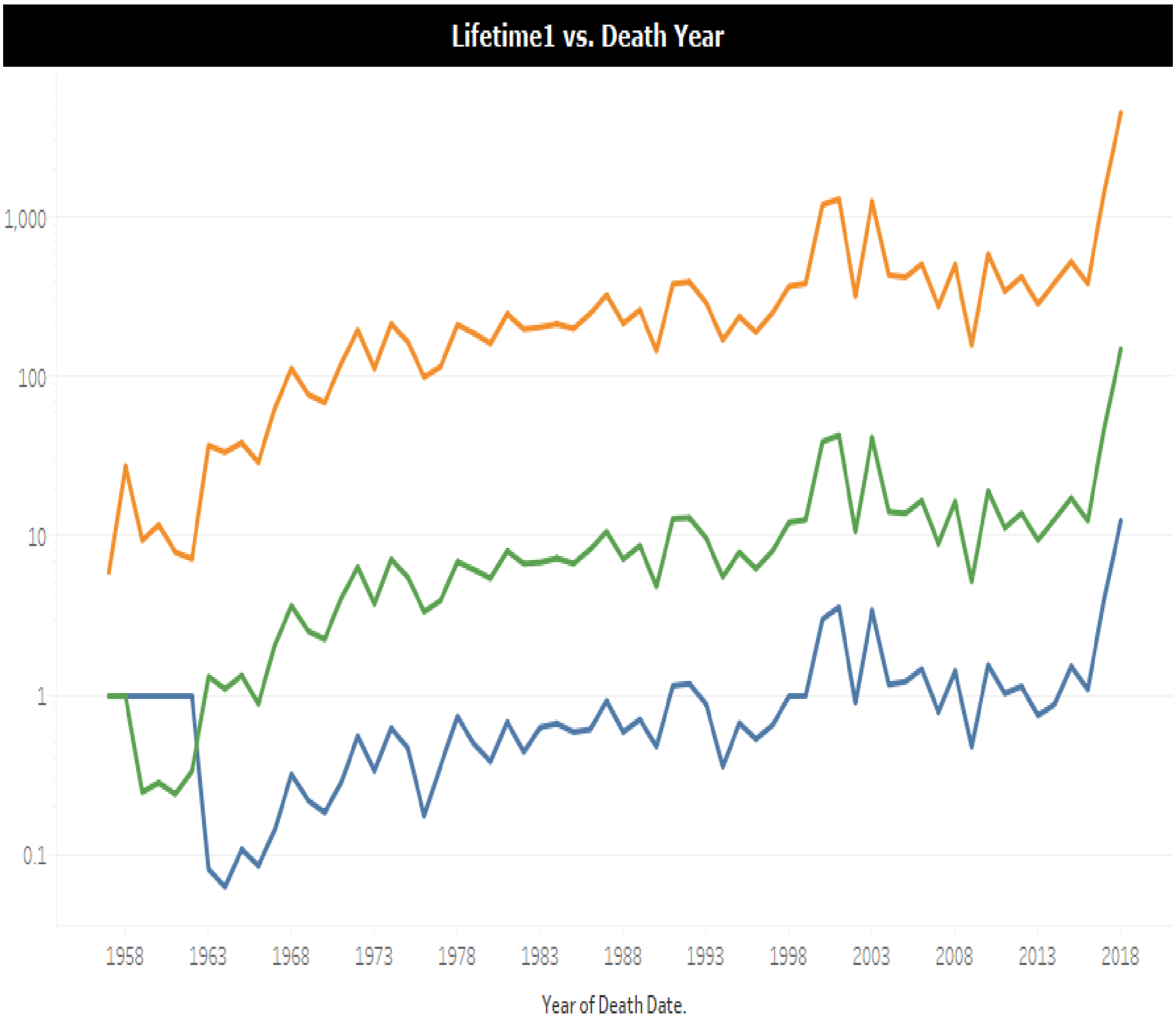

Figure 32: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime1 in Days, Months and Years Vs. Death Year. X Axis Represents Year of Death Date. Y Axis Represents Average Lifetime1 in Days, Months and Years. The early parts of the blue and green curves plot as 1 but are really 0. Orange: Lifetime in Days, Green: Lifetime in Months, Blue: Lifetime in Years.

Figure 32 shows all charts in one place for simple comparison analysis. The average lifetimes in days, months and years were increasing for a while and thereafter Lifetime1 fluctuates. Lifetime1 in days, months and years followed the same pattern with respect to Death Year.

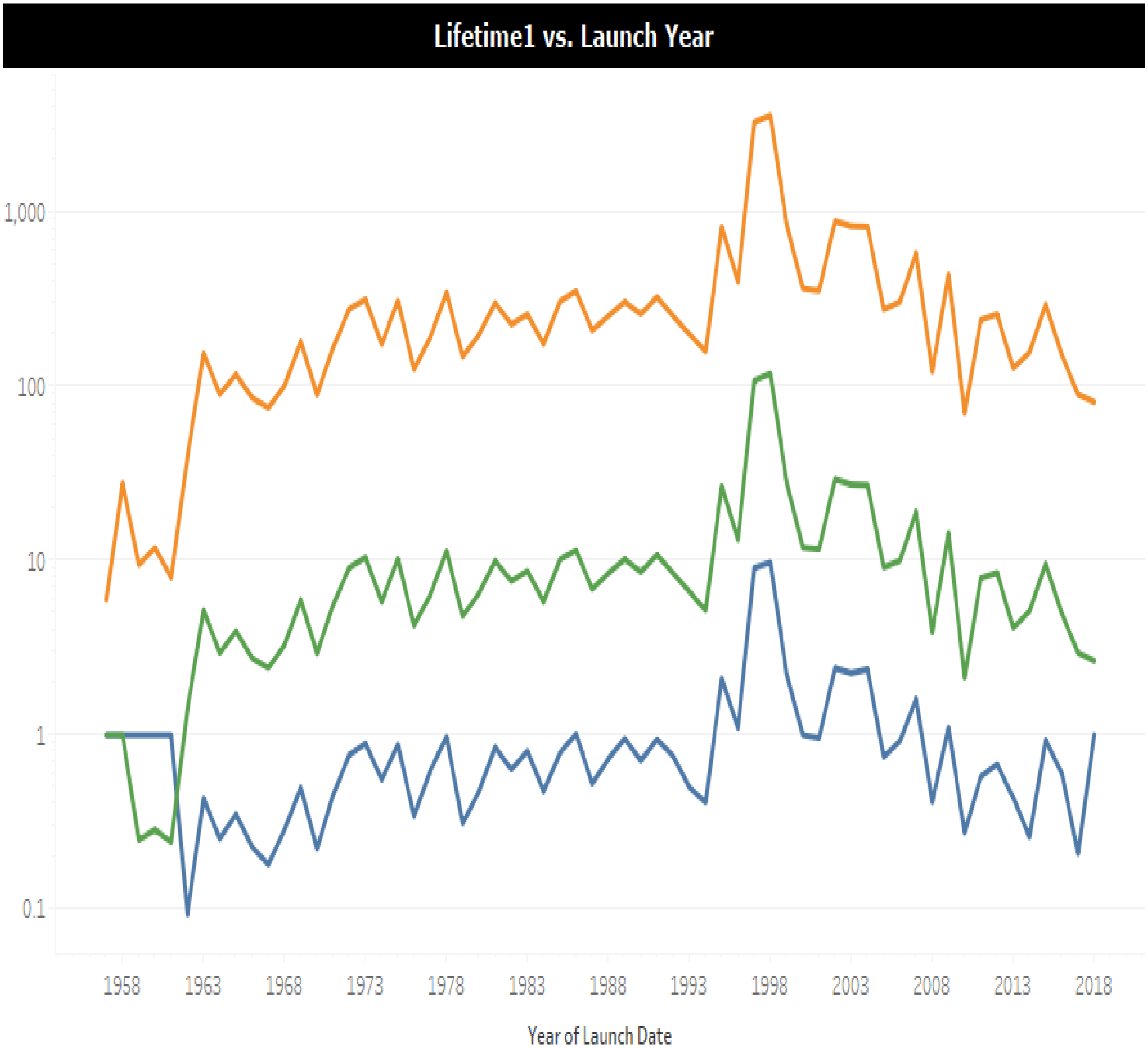

Figure 33: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime1 in Days, Months and Years Vs. Launch Year. X Axis Represents Year of Launch Date. Y Axis Represents Average Lifetime1 in Days, Months and Years. The early parts of the blue and green curves plot as 1 but are really 0. Orange: Lifetime in Days, Green: Lifetime in Months, Blue: Lifetime in Years

In Figure 33, averages for Lifetime1 in Days, Months and Years with respect to Launch Year are shown. The average lifetime was gradually increasing and then gradually decreasing. The three conditions show a similar pattern.

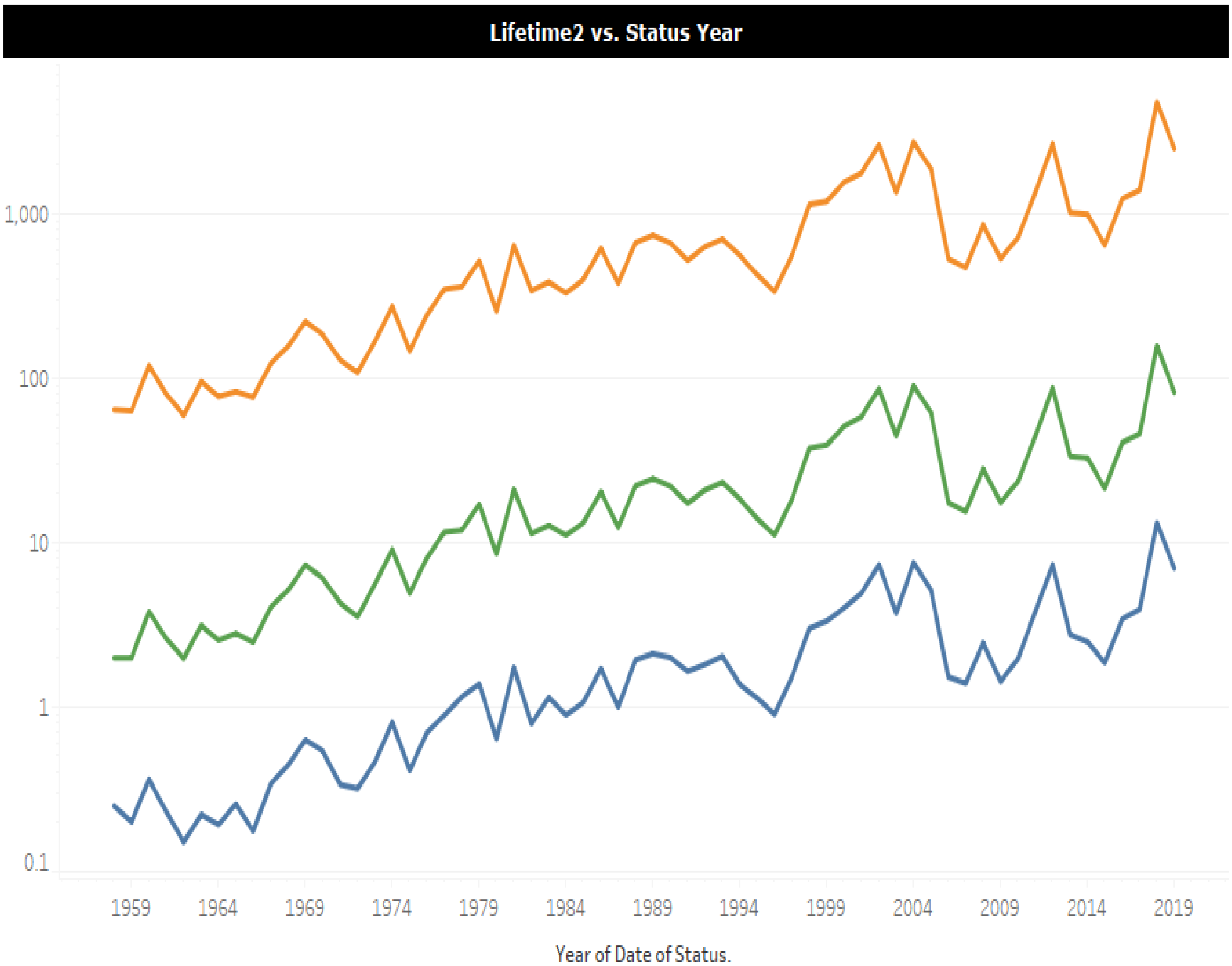

Figure 34: New Death Data Satellites with Applicable Status Dates in SATCAT: Average Lifetime2 in Days, Months and Years vs. Status Year. X Axis Represents Year of Status Date. Y Axis Represents Average Lifetime2 in Days, Months and Years. Orange: Lifetime in Days, Green: Lifetime in Months, Blue: Lifetime in Years

Figure 33 and Figure 34 show the relation between averages of Lifetime1 with respect to Launch Year and averages of Lifetime2 with respect to Status Year.

From this analysis I considered reasons for the sudden changes or fluctuations in averages of lifetimes of satellites. Here are some possible reasons,

- Random noise

- External events like politics, recessions etc.
- 10% of big number is a bigger fluctuation than 10% of a smaller number

6. New Death Data Without Kosmos and with Only Kosmos

From New Death Data records (5610) I separated the Kosmos satellites data. The remaining records total to 3392. The Kosmos records total to 2218. From this data, I analyzed the behavior of yearly averages of lifetimes with respect to Death Year.

6.1 New Death Data Without Kosmos

New Death Data without Kosmos satellites total 3392 records. Lifetime1 was calculated from Death date and Launch Date. The comparison analysis is shown in the following figures.

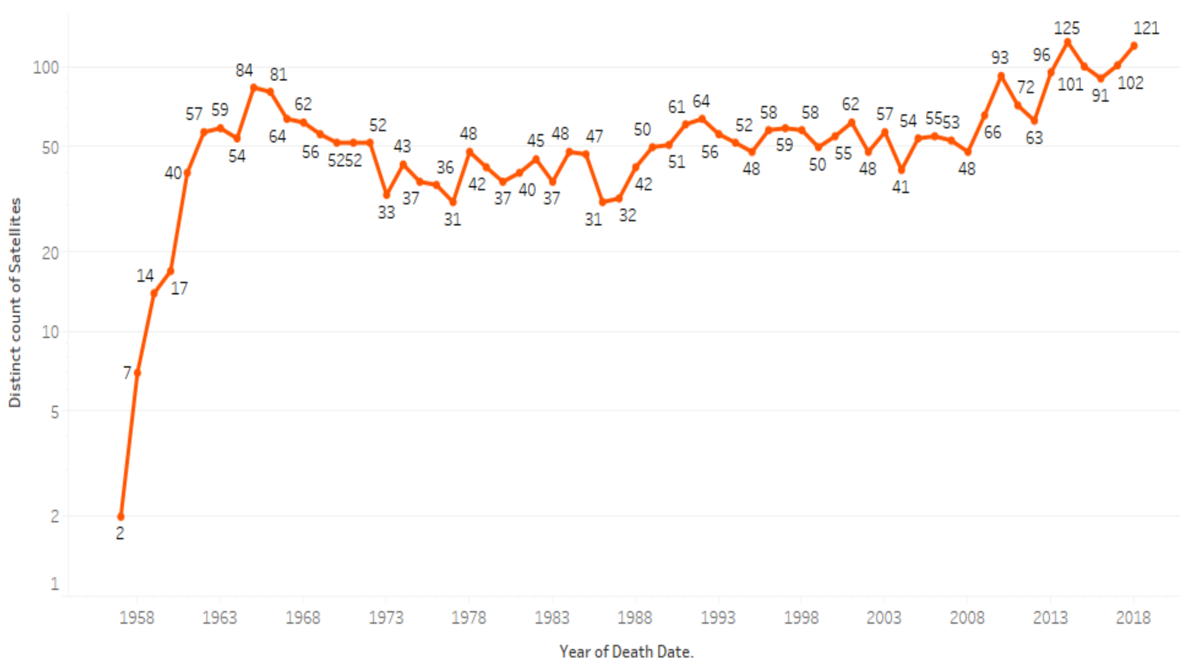

Figure 35: New Death Data Without Kosmos: Distinct Count of Satellites Vs. Death Year.

Figure 35 shows the Distinct count of satellites vs. Death Year. Satellite count increased rapidly at first and then it gradually decreased until 1990 and thereafter it went up gradually.

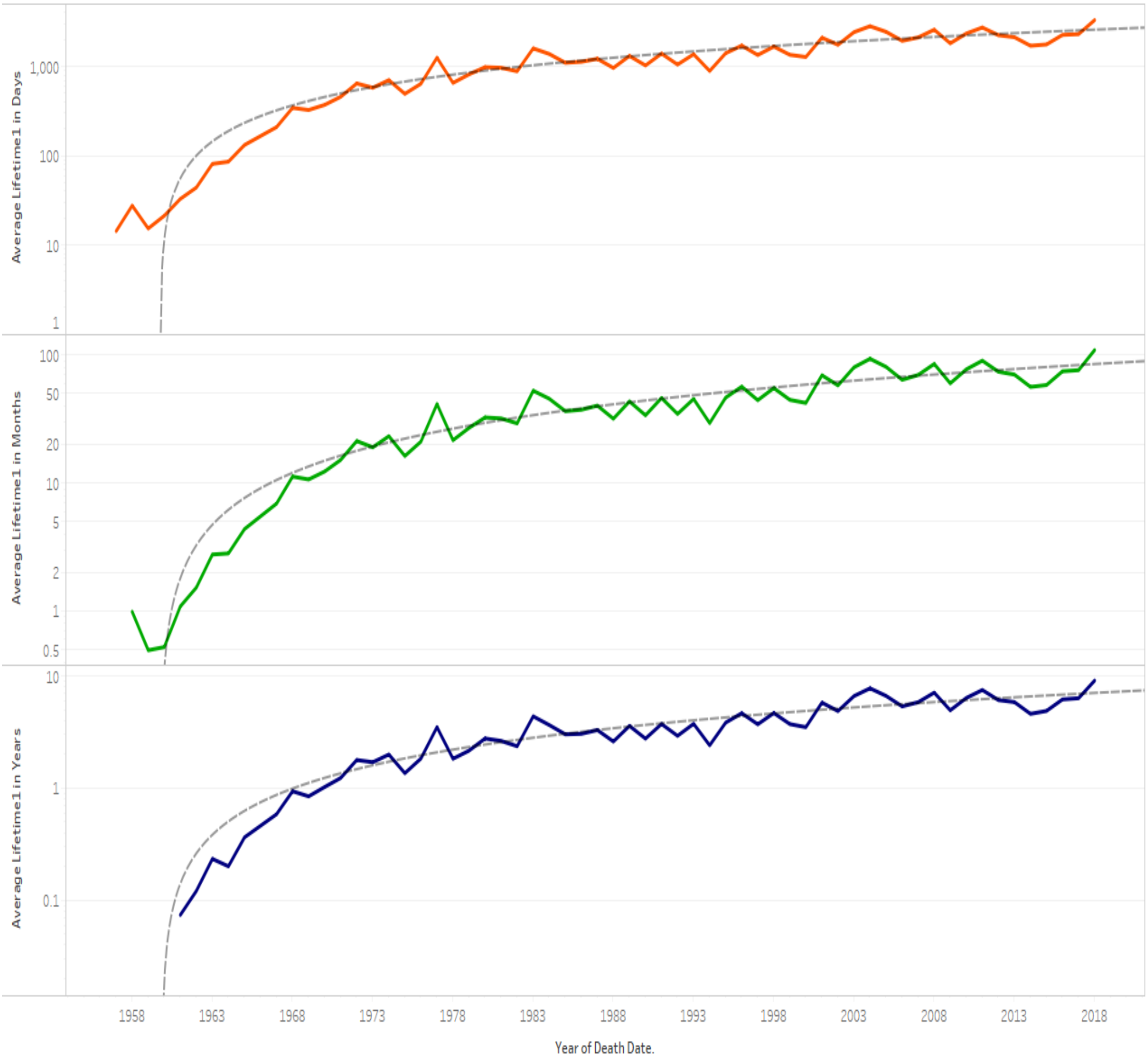

Figure 36: New Death Data Without Kosmos Satellites: Average Lifetimes in Days, Months and Years Vs. Death Year.

Orange: Lifetime in Days, Green: Lifetime in Months, Blue: Lifetime in Years

Figure 36 shows the average of Lifetime1 in Days, Months and Years vs. Death Year. Lifetime1 increased rapidly and at start Lifetime1 has rapid growth. All lifetimes whether measured in Days, Months or Years followed the same pattern.

6.2 New Death Data with Only Kosmos

In New Death Data for only Kosmos satellites the total number of records is 2218. The data was analyzed in Tableau.

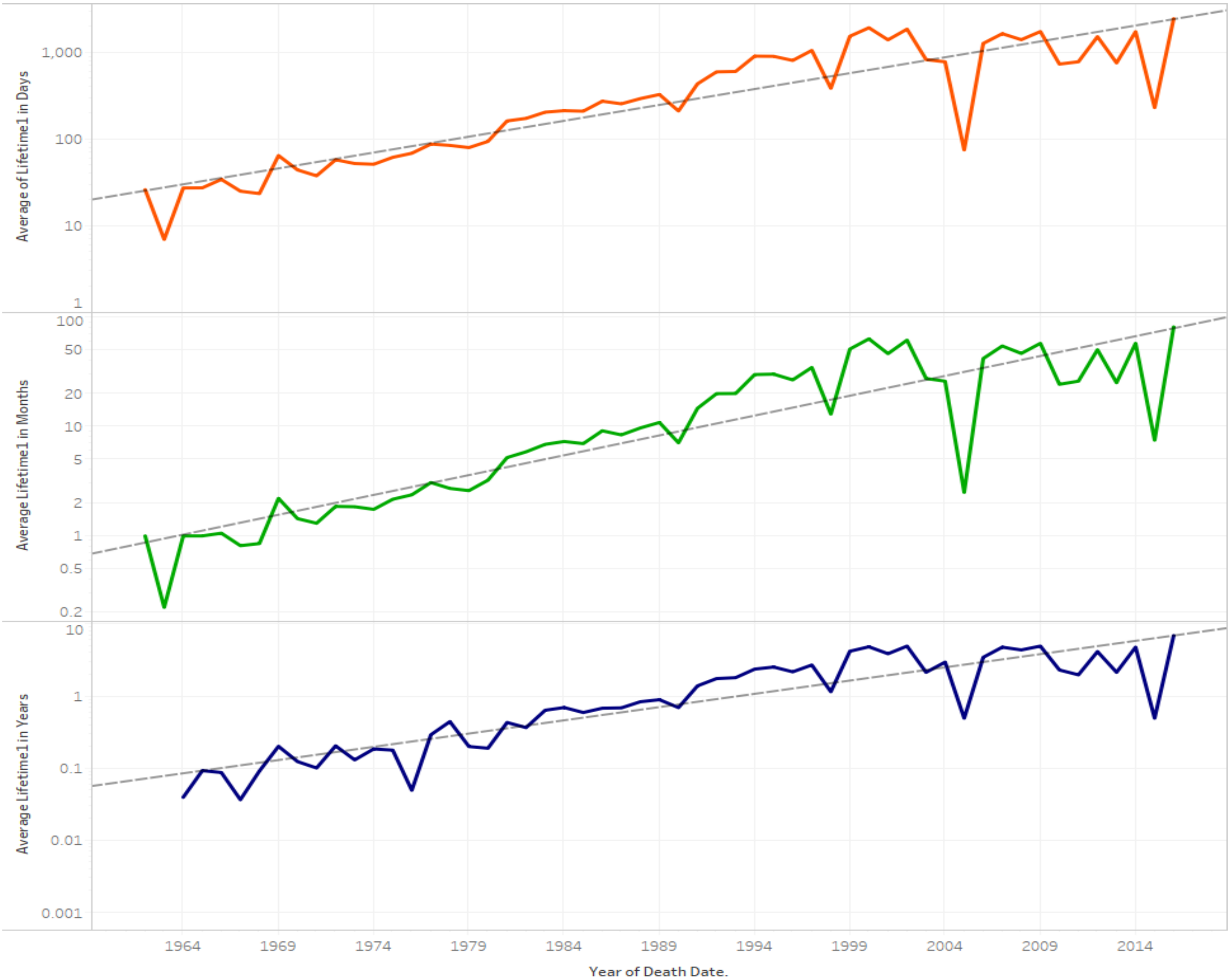

Figure 37: New Death Data for Kosmos Satellites: Average lifetimes in Days, Months and Years vs. Death Year. Orange: Lifetime in Days, Green: Lifetime in Months, Blue: Lifetime in Years

Figure 37 shows the averages of Lifetime1 in Days, Months and Years vs. Death Year. From the above chart, I observed that Lifetime1 increased linearly on plots with the Y-Axis scaled logarithmically indicating exponential increase. All lifetimes whether in Days, Months and Years followed the same pattern.

Conclusion

This project focused on yearly averages for lifetimes of Earth satellites. The lifetimes were calculated based on Death Date and Status Date. Recall that if “In Earth Orbit” statuses are filtered out, as done here, according to McDowell (personal communication), "What you call "Date of Status" or "SATCAT status date" is better termed "decay date" or "reentry date". It's the date the satellite reentered or was deorbited. The lifespan increased almost exponentially with respect to Death Date but not exactly.

From the comparison of average lifetime with respect to Death Date and Status Date vs. launch date, more meaningful insights were found with average lifetime using Death Date and not so much using Status Year-based lifetimes. Distinct counts of satellites with Death Dates and Status Dates vs. Launch Date followed similar patterns in all conditions.

From this analysis, I observed that Lifetime1 was reliable for all 3 conditions: Death Year, Launch Year and Status Year. Lifetime2 was reliable only with Status Year. Lifetime1 and Lifetime 2 have the same patterns with respect to Status Year but not with Death Year. Finally, Lifetime1 and Lifetime2 have a bigger difference in average of lifetimes of satellites during the start period of satellite technology and thereafter the difference decreased.

Measuring lifespans with a granularity of days provides more accurate results than using a granularity of months or years. However, the three results are generally similar.

Plotting results on the X axes for launch date vs. status date vs. death date, I noticed that there are more insights possible in the analysis and in future work I am going to focus on how Moore's law and Wright's law relate to the average lifespans.

References

1. J. McDowell, Jonathan's space report, planet4589.org; Guide to the SATACT columns: SATCAT, https://planet4589.org/space/log/satcat.html

2. J. Gonzalo, D. Dominguez and D. López, "On the challenge of a century lifespan satellite", Progress in Aerospace Sciences, 2014, vol. 70, pp. 28–41.
3. V. Kodali, R. Duggirala, R.S. Segall, H. Aboudja, and D. Berleant, "Travel to extraterrestrial bodies over time: some exploratory analyses of mission data", (2018) Journal of the Arkansas Academy of Science: Vol. 72, Article 13.
4. D. Berleant, V. Kodali, R.S. Segall, H. Aboudja, and M. Howell, "Moore's law, Wright's law and the countdown to exponential space", The Space Review, Jan. 7, 2019, http://thespacereview.com/article/3632/1
5. SpacePunx, Space Population.png, Population of space over time,2016.
6. List of satellites, en.wikipedia.org/wiki/Category: Lists of satellites
7. V. Kodali, "Earth Satellite Data Analysis", Master's Thesis, May-2018, UALR
8. V. Kodali, "Planet 4589 Update", Unpublished Document, 2019.
9. UCS Satellite Database, https://www.ucsusa.org/resources/satellite-database
10. Introduction to Tableau, https://intellipaat.com/blog/what-is-tableau/
11. Satellites Data, https://space.skyrocket.de/
12. Introduction to Power BI, https://www.tutorialspoint.com/power_bi/power_bi_introduction.htm
13. List of Kosmos Satellites, https://en.wikipedia.org/wiki/List_of_Kosmos_satellites
14. Global Space Industry Outlook Report 2019-2025: Satellite Manufacturing, Launch Services, Downstream, Ground Stations and Satellite Networks, Value Chain, https://www.prnewswire.com/news-releases

Appendix

Python code for Data Cleaning and Merging

```
In [1]: # Loading Libraries

import os
```

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

os.chdir("C:\\Users\\jaipa\\Desktop\\H")

In [2]: ## Reading-in New-Death-data

Death_dates=pd.read_excel('Death Dates.xlsx')

Death_dates.head()

In [3]: Death_dates.shape

In [4]: Death_dates['Satellite_Catalog'].nunique()

In [5]: Death_dates.isnull().sum()

Out[2]:

Satellite_Catalog Death Date Death_Year Death_Month Death_Day

0 S000002 1957 Oct 27 1957 Oct 27

1 S000003 1957 Nov 9 1957 Nov 9

2 S000004 1958 Apr 1 1958 Apr 1

3 S000005 1964 Mar 1964 Mar NaN

4 S000006 1958 May 10 1958 May 10

Out[3]: (9031, 5)

Out[4]: 9031

Out[5]: Satellite_Catalog 0

Death Date 0

Death_Year 0
```

Death_Month 3325

Death_Day 4186

dtype: int64

9/25/2019 Complete Analysis 1 _ Print

localhost:8888/nbconvert/html/Complete Analysis 1 _ Print.ipynb?download=false 2/7

In [6]: *## Eliminating rows having Death_Month as 'U','E','?'*

```
Death_dates_1=Death_dates.loc[(Death_dates['Death_Month'].notnull())&(~Death_d
ates['Death_Month'].isin(['U','E','?'])),:]
Death_dates_1.head()
```

In [7]: *## Stripping extra spaces*

```
Death_dates_1['Death_Month']=Death_dates_1['Death_Month'].str.strip()
```

In [8]: *## Stripping Special characters*

```
def strip_special(special):
return special.replace("?","").replace("Arp",'Apr')
```

In [9]: Death_dates_1['Death_Month']=Death_dates_1['Death_Month'].apply(strip_special)

In [10]: Death_dates_1['Death_Month'].value_counts()

In [11]: Death_dates_1.isnull().sum()

In [13]: Death_dates_1['Death_Day']=Death_dates_1['Death_Day'].astype('str')

In [14]: Death_dates_1['Death_Day']=Death_dates_1['Death_Day'].str.strip()

Out[6]:

Satellite_Catalog Death Date Death_Year Death_Month Death_Day

0 S000002 1957 Oct 27 1957 Oct 27

1 S000003 1957 Nov 9 1957 Nov 9

2 S000004 1958 Apr 1 1958 Apr 1

3 S000005 1964 Mar 1964 Mar NaN

4 S000006 1958 May 10 1958 May 10

Out[10]: Dec 551

Oct 544

Sep 526

Feb 479

Jun 478

Nov 463

Aug 455

May 454

Apr 451

Mar 435

Jul 421

Jan 355

Name: Death_Month, dtype: int64

Out[11]: Satellite_Catalog 0

Death Date 0

Death_Year 0

Death_Month 0

Death_Day 767

dtype: int64

In [15]: Death_dates_1['Death_Day']=Death_dates_1['Death_Day'].apply(strip_special)

In [17]: *# Converting to date-time format*

Death_dates_1['Death_Day']=pd.to_numeric(Death_dates_1['Death_Day'],errors='co

```
erce')

In [18]: Death_dates_1['Death_Day'].isnull().sum()

In [19]: # Reading in Sat-Cat Data

Sat_cat=pd.read_excel('Sat_cat_Autoupdate.xlsx')

Sat_cat.head()

In [20]: # Stripping spaces

Sat_cat['Satellite_Catalog']=Sat_cat['Satellite_Catalog'].str.strip()

In [21]: Sat_cat['Satellite_Catalog'].dtypes

In [22]: Death_dates_1.loc[4010,'Satellite_Catalog']

In [23]: Sat_cat.loc[4010,'Satellite_Catalog']

In [24]: Death_dates_1['Satellite_Catalog']=Death_dates_1['Satellite_Catalog'].str.uppe

r()

Out[18]: 767

In [25]: Death_dates_1['Satellite_Catalog'].isin(Sat_cat['Satellite_Catalog']).value_co

unts()

In [26]: # Merging Satcat data and Death_Dates data

df=pd.merge(Sat_cat,Death_dates_1,on='Satellite_Catalog')

df.head()

In [27]: df.shape

In [28]: df=df.loc[df['Launch_Date'].notnull(),:]

In [29]: df['Launch_Date']=pd.to_datetime(df['Launch_Date'])

In [30]: import calendar

df['Launch_Month']=df['Launch_Date'].dt.month

df['Launch_Month']=df['Launch_Month'].astype('int')

In [31]: df['Launch_Month']=df['Launch_Month'].apply(lambda x: calendar.month_abbr[x])
```

Out[25]: True 5612

Name: Satellite_Catalog, dtype: int64

Out[27]: (5612, 15)

9/25/2019 Complete Analysis 1 _ Print

localhost:8888/nbconvert/html/Complete Analysis 1 _ Print.ipynb?download=false 5/7

In [32]: *# Imputing Day value in Death Dates*

```
M=[]
for index,row in df.iterrows():
if pd.isnull(row['Death_Day'])==True:
if (row['Launch_Date'].year == row['Death_Year'])&(row['Launch_Month']
==row['Death_Month']):
M.append(round((30-row['Launch_Date'].day)/2))
else:
M.append(15)
else:
M.append(row['Death_Day'])
```

In [33]: df['Death_Day_1']=M

In [34]: df['Death_Day_1']=df['Death_Day_1'].astype('int')

In [66]: df=df.loc[df['Death_Day_1']!=0,:]

In [67]: df['Death_Date_1']=df['Death_Year'].astype('str')+'/'+df['Death_Month']+'/'+df

['Death_Day_1'].astype('str')

In [68]: df.loc[1,'Death_Date_1']

In [69]: from datetime import datetime

In [74]: N=[]

```
for index,row in df.iterrows():
```

N.append(datetime.strptime(row['Death_Date_1'], '%Y/%b/%d'))

In [76]: df['Death_Date_1']=N

Out[68]: '1957/Nov/9'

9/25/2019 Complete Analysis 1 _ Print

localhost:8888/nbconvert/html/Complete Analysis 1 _ Print.ipynb?download=false 6/7

In [77]: df.head()

In [78]: df.to_csv('Test_11.csv',index=**False**)

In [59]: print(datetime.strptime(df.loc[1,'Death_Date_1'], '%Y/%b/%d'))

In [60]: datetime.strptime(df.loc[1,'Death_Date_1'], '%A, %B %d, %Y')

In [36]: df.head()