Meteor Fall Analysis



Index

1. Introduction………………………………………………………………………………………………………………….3
2. Context…………………………………………………………………………………………………………………………3
3. Complication…………………………………………………………………………………………………………………3
4. Dataset ………………………………………………………………………………………………………………………..3
5. Analysis of Dataset……………………………………………………………………………………………………….4
6. Data cleansing………………………………………………………………………………………………………………4
7. Boot strapping……………………………………………………………………………………………………………..5
8. Visualization…………………………………………………………………………………………………………………10
9. Conclusion……………………………………………………………………………………………………………………16
10. **Introduction:**

Meteorite falls, also called observed falls, are meteorites collected after their fall from space was observed by people or automated devices. All other meteorites are called "finds". There are more than 1,100 documented falls listed in widely used databases, most of which have specimens in modern collections. As of July 2016, the Meteoritical Bulletin Database has 1,147 confirmed falls (excluding rocks found nearby on the ground which turn out to not be associated with the fall and those with doubtful status).

Material from observed falls has not been subjected to terrestrial weathering, making the find a better candidate for scientific study. Historically, observed falls were the most compelling evidence supporting the extraterrestrial origin of meteorites. Furthermore, observed fall discoveries are a better representative sample of the types of meteorites which fall to Earth. For example, iron meteorites take much longer to [weather](https://en.wikipedia.org/wiki/Meteorite_weathering) and are easier to identify as unusual objects, as compared to other types. This may explain the increased proportion of iron meteorites among finds (6.7%), over that among observed falls . There is also detailed [statistics on falls](https://en.wikipedia.org/wiki/Meteorite_fall_statistics) such as based on [meteorite classification](https://en.wikipedia.org/wiki/Meteorite_classification).

1. **Context:**

Meteorites can belong to different classes and can have different masses. Meteorites with less mass can be less dangerous when compared to meteorites with huge mass. So, we have tried analyzing the dataset with considerable size and calculated the average mass of the meteorite. Now we are trying to find out the average mass of meteorite that can strike the earth for the coming 10 years till 2024. We would also like to find out the location of the meteorite in INDIA based on the geo codes. We would like to provide an estimate on where the Meteor would be fall in the coming in the coming Ten years giving a radius of the distance based on meteorite fall in the recent years. Now again, considering INDIA, we would try and analyze the pattern of the location movement of the Meteorite fall in next 10 years.

1. **Complication:**

When we converted the geo codes available and into a world data that was available for geo codes, we saw that a lot of locations we missing and hence we used revgeococde available to convert the geo code into locations. We manually calculated the geocodes of INDIA . We have made a subset of this country to analyze the data.

1. **Dataset:**

We have obtained the dataset Meteorite Landings from kaggle(<https://www.kaggle.com/nasa/meteorite-landings> ). It contains a data of 31k rows with data from 860 AD till 2016 AD for all the locations with different types of meteorites and their names and geo locations.

1. **Analysis of Dataset:**

Different Attributes of the dataset are given as below

1. **name**: the name of the meteorite (typically a location, often modified with a number, year, composition, etc)
2. **id**: a unique identifier for the meteorite
3. **nametype**: one of:   
   -- valid: a typical meteorite   
   -- relict: a meteorite that has been highly degraded by weather on Earth
4. **recclass**: the class of the meteorite; one of a large number of classes based on physical, chemical, and other characteristics .
5. **mass**: the mass of the meteorite, in grams
6. **fall**: whether the meteorite was seen falling, or was discovered after its impact; one of:   
   -- Fell: the meteorite's fall was observed   
   -- Found: the meteorite's fall was not observed
7. **year**: the year the meteorite fell, or the year it was found (depending on the value of fell)
8. **reclat**: the latitude of the meteorite's landing
9. **reclong**: the longitude of the meteorite's landing
10. **GeoLocation**: a parentheses-enclose, comma-separated tuple that combines reclat and reclong.
11. **Data Cleansing:**

Code for cleansing the data is given below.

#loading the required packages

library(ggplot2)

library(ggmap)

library(xlsx)

library(tcltk)

library(sqldf)

library(pastecs)

## loading the data file

meteriod <- read.csv(file.choose(), sep = ",", header = TRUE)

head(meteriod)

### cleaning the dataset by selecting the years between 860 and 2016

attach(meteriod)

meteriodt1 <- subset(meteriod, (year >=860 & year <= 2016) )

meteriodt1 <- subset(meteriodt1,(reclong<=180 & reclong>=-180 & (reclat!=0 | reclong!=0)))

meteriodt1 <- na.omit(meteriodt1)

# creating a sample data.frame with your lat/lon points

longlat <- as.data.frame(cbind(meteriodt1$reclong,meteriodt1$reclat))

## for fetching the location based on the longitude and latitude - but not working

for (i in 20001:nrow(meteriodt1))

{

meteriodt1$location[i]<- revgeocode(c(meteriodt1$reclong[i], meteriodt1$reclat[i]))

}

#### longitude and latitude and region world

world <- map\_data("world")

head(world)

## plotting the meteroid locations

ggplot() +

theme\_bw() +

geom\_polygon(data=world, aes(x=long, y=lat, group=group), fill=NA, color="blue") +

labs(y="latitude", x="longitude", alpha="Density of meteorites", title=paste("Density of all meteorite impacts in the world")) +

stat\_density2d(na.rm=TRUE,data=meteriodt1, aes(x=reclong, y=reclat, alpha=..level..),geom="polygon",fill="#FF2222")

# Antartica longitude, latitude and region of the world

worldantartic <- subset(world, (world$region == "Antarctica"))

worldantartic <- worldantartic[order(worldantartic$lat),]

# Antarctic meteorite finds are so common that The Meteoritical Society (the source of this dataset)

# has an option to only search non-Antarctic meteorites, just so the Antarctic finds don't completely overwhelm your search results.

ggplot(meteriodt1, aes(x=reclat)) +

geom\_histogram(binwidth=5) +

scale\_x\_continuous(breaks=seq(-90,90,by=10)) +

labs(y="count", x="lattitude", title=paste("Plot of count of all meteorite impacts based on the lattitude"))

## Non Antartica longitude, latitude and region of the world

worldt <- subset(world, (world$lat <= -70))

antarticlatstarts <- max(worldantartic$lat)

## with out ntartica regions metroid data

meteroidnonantartic <- subset(meteriodt1, (meteriodt1$reclat >= antarticlatstarts))

ggplot() +

theme\_bw() +

geom\_polygon(data=world, aes(x=long, y=lat, group=group), fill=NA, color="blue") +

labs(y="latitude", x="longitude", alpha="Density of meteorites", title=paste("Density of all non meteorite impacts in the world")) +

stat\_density2d(na.rm=TRUE,data=meteroidnonantartic, aes(x=reclong, y=reclat, alpha=..level..),geom="polygon",fill="#FF2222")

Now we have cleansed all the data that was taken from Kaggle.

**Boot strapping:**

1. to find out the average mass of meteorite that can strike the earth for the coming 10 years till 2024:

Meteorites can belong to different classes and can have different masses. Meteorites with less mass can be less dangerous when compared to meteorites with huge mass. So, we have tried analyzing the dataset with considerable size and calculated the average mass of the meteorite. Now we are trying to find out the average mass of meteorite that can strike the earth for the coming 10 years till 2024.

## fetching fell details, means removing the found details

meteriodt2 <- meteriodt1

meteriodt2 <- subset(meteriodt2, (meteriodt2$fall == "Fell" ))

## order by year

meteriodt2 <- meteriodt2[order(meteriodt2$year),]

## year, cnt, sum(mass), avg(mass)

metetroidyearstat <- sqldf("Select year, count(\*), SUM(mass) as sumofmass, AVG(MASS) as avgmass from meteriodt2 group by year")

metetroidyearstat$avgchangeinmass[1] <- 0

for( i in 2: dim(metetroidyearstat)[1])

{

metetroidyearstat$avgchangeinmass[i] <- (metetroidyearstat$avgmass[i] - metetroidyearstat$avgmass[i-1]) / (metetroidyearstat$avgmass[i-1])

}

## performing bootstrap for the stats dataset taking the year above 1900 and for 3 years and performing bootstrap for 1000 simulations

metetroidyearstat <- subset(metetroidyearstat, (metetroidyearstat$year >= 1900))

metetroidyearstat <- metetroidyearstat[-114,]

## for 3 years taking initial change of 2014 as 2013's

for (i in 1:3)

{

year <- metetroidyearstat$year[dim(metetroidyearstat)[1]] + i

cnt <- 0

sumofmass <- 0

avgmass <- metetroidyearstat$avgmass[dim(metetroidyearstat)[1]] #taking 2013 avgmass

randomchnge <- runif(1, min(metetroidyearstat$avgchangeinmass), max(metetroidyearstat$avgchangeinmass))

print(randomchnge)

avgmass <- avgmass\* (1+randomchnge)

metetroidyearstattest <- data.frame(year,cnt,sumofmass,avgmass,randomchnge)

colnames(metetroidyearstattest) <- colnames(metetroidyearstat)

metetroidyearstat <- rbind(metetroidyearstat,metetroidyearstattest)

}

## taking median of the avgmass change

for (i in 1:3)

{

year <- metetroidyearstat$year[dim(metetroidyearstat)[1]] + 1

cnt <- 0

sumofmass <- 0

avgmass <- median(metetroidyearstat$avgmass) #taking median of the avgmass change

randomchnge <- runif(1, min(metetroidyearstat$avgchangeinmass), max(metetroidyearstat$avgchangeinmass))

print(randomchnge)

avgmass <- avgmass\* (1+randomchnge)

metetroidyearstattest <- data.frame(year,cnt,sumofmass,avgmass,randomchnge)

colnames(metetroidyearstattest) <- colnames(metetroidyearstat)

metetroidyearstat <- rbind(metetroidyearstat,metetroidyearstattest)

}

## performing bootstrap for 1000 simulations

bootstrap <- data.frame()

for(j in 1:1000)

{

metetroidyearstat <- metetroidyearstat[-114:-123,]

## taking median of the avgmass chang

for (i in 1:10)

{

year <- metetroidyearstat$year[dim(metetroidyearstat)[1]] + 1

cnt <- 0

sumofmass <- 0

avgmass <- median(metetroidyearstat$avgmass) #taking median of the avgmass change

randomchnge <- runif(1, min(metetroidyearstat$avgchangeinmass), max(metetroidyearstat$avgchangeinmass))

print(randomchnge)

avgmass <- avgmass\* (1+randomchnge)

metetroidyearstattest <- data.frame(year,cnt,sumofmass,avgmass,randomchnge)

colnames(metetroidyearstattest) <- colnames(metetroidyearstat)

metetroidyearstat <- rbind(metetroidyearstat,metetroidyearstattest)

}

final2024mass <- metetroidyearstat$avgmass[123]

bootstraptest <- data.frame(j,final2024mass)

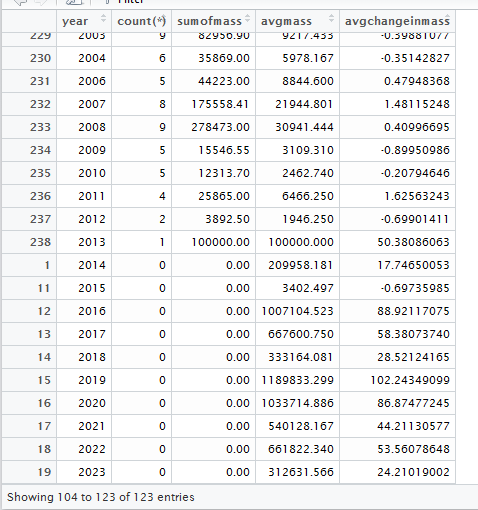
bootstrap <- rbind(bootstrap,bootstraptest)

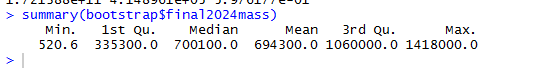
}

head(bootstrap)

stat.desc(bootstrap$final2024mass)

summary(bootstrap$final2024mass)





1. Bootstrapping for average distance change between the locations:

We would also like to find out the location of the meteorite in INDIA based on the geo codes. We would like to provide an estimate on where the Meteor would be fall in the coming in the coming Ten years giving a radius of the distance based on meteorite fall in the recent years.

## performing bootstrap for 1000 simulations order us details based on year, lat, long

meteroidust1 <- meteroidus

meteroidust1$distchange[1] <- 0

meteroidust1$distchange[2] <- 0

## finding dist change

for( i in 3: dim(meteroidust1)[1])

{

meteroidust1$distchange[i]<- (meteroidust1$dist[i] - meteroidust1$dist[i-1]) / (meteroidust1$dist[i-1])

}

plot(meteroidust1$year, meteroidust1$distchange)

bootstrap2 <- data.frame()

for(j in 1:1000)

{

meteroidust1 <- meteroidust1[-162:-171,]

## taking median of the avgmass chang

for (i in 1:10)

{

year <- meteroidust1$year[dim(meteroidust1)[1]] + 1

reclat <- 0

reclong <- 0

dist <- median(meteroidust1$dist)

randomchnge <- runif(1, min(meteroidust1$distchange), max(meteroidust1$distchange))

print(paste(dist, randomchnge,randomchnge/10000, sep = " "))

dist <- dist\* (1+ (randomchnge/100) )

meteroidust1test <- data.frame(year,reclat,reclong,dist,randomchnge)

colnames(meteroidust1test) <- colnames(meteroidust1)

meteroidust1 <- rbind(meteroidust1,meteroidust1test)

}

final2021dist <- meteroidust1$dist[171]

initialchange <- meteroidust1$distchange[dim(meteroidust1)[1] -10]

finalchange <- meteroidust1$distchange[dim(meteroidust1)[1]]

changeperct <- ((finalchange - initialchange)/initialchange)

bootstraptest <- data.frame(j,changeperct,final2021dist)

bootstrap2 <- rbind(bootstrap2,bootstraptest)

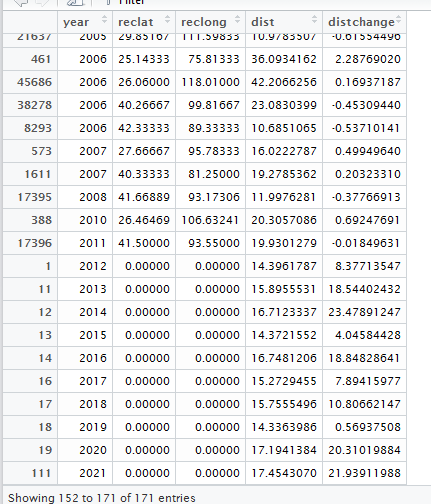
}

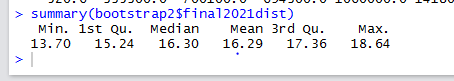
stat.desc(bootstrap2$changeperct)

summary(bootstrap2$final2021dist)

head(meteroidust1)

write.xlsx(meteroidust1, "A:/meteroidust1.xlsx")





1. Analyzing the pattern of the location of geocodes:

Now again, considering INDIA, we would try and analyze the pattern of the location movement of the Meteorite fall in next 10 years.

Code for bootstrapping by average change of location codes:

## performing bootstrap for 1000 simulations here ordered us details based on lat, long

meteroidust2 <- meteroidus

## removing year

meteroidust2 <- meteroidust2[,-1]

## order by

meteroidust2 <- meteroidust2[order(meteroidust2$reclat,meteroidust2$reclong),]

meteroidust2$distchange[1] <- 0

meteroidust2$distchange[2] <- 0

## finding dist change

for( i in 3: dim(meteroidust2)[1])

{

meteroidust2$distchange[i]<- (meteroidust2$dist[i] - meteroidust2$dist[i-1]) / (meteroidust2$dist[i-1])

}

head(meteroidust2)

bootstrap3 <- data.frame()

meteroidust2 <- subset(meteroidust2, (meteroidust2$reclat != 31.20000))

for(j in 1:1000)

{

meteroidust2 <- meteroidust2[-161:-170,]

## taking median of the avgmass chang

for (i in 1:10)

{

reclat <- 0

reclong <- 0

dist <- median(meteroidust2$dist) #taking median of the dist

randomchnge <- runif(1, min(meteroidust2$distchange), max(meteroidust2$distchange))

print(paste(dist, randomchnge, randomchnge/10000, sep = " "))

dist <- dist\* (1+(randomchnge/1000) )

meteroidust2test <- data.frame(reclat,reclong,dist,randomchnge)

colnames(meteroidust2test) <- colnames(meteroidust2)

meteroidust2 <- rbind(meteroidust2,meteroidust2test)

}

final2021dist <- meteroidust2$dist[170]

initialchange <- meteroidust2$distchange[dim(meteroidust2)[1] -10]

finalchange <- meteroidust2$distchange[dim(meteroidust2)[1]]

changeperct <- ((finalchange - initialchange)/initialchange)

bootstraptest <- data.frame(j,changeperct,final2021dist)

bootstrap3 <- rbind(bootstrap3,bootstraptest)

}

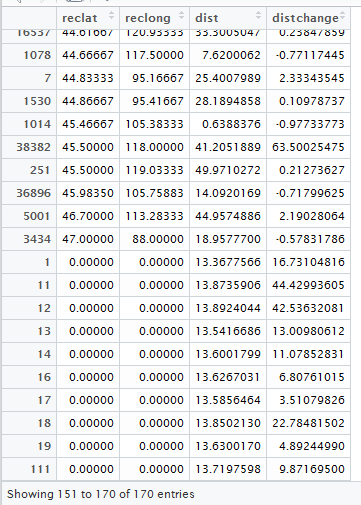
stat.desc(bootstrap3$changeperct)

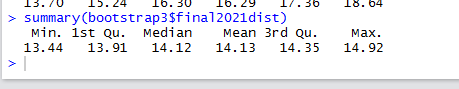
summary(bootstrap3$final2021dist)

summary(bootstrap3$final2021dist)

head(meteroidust2)

write.xlsx(meteroidust2, "A:/meteroidust2.xlsx")



****

**Visualization:**

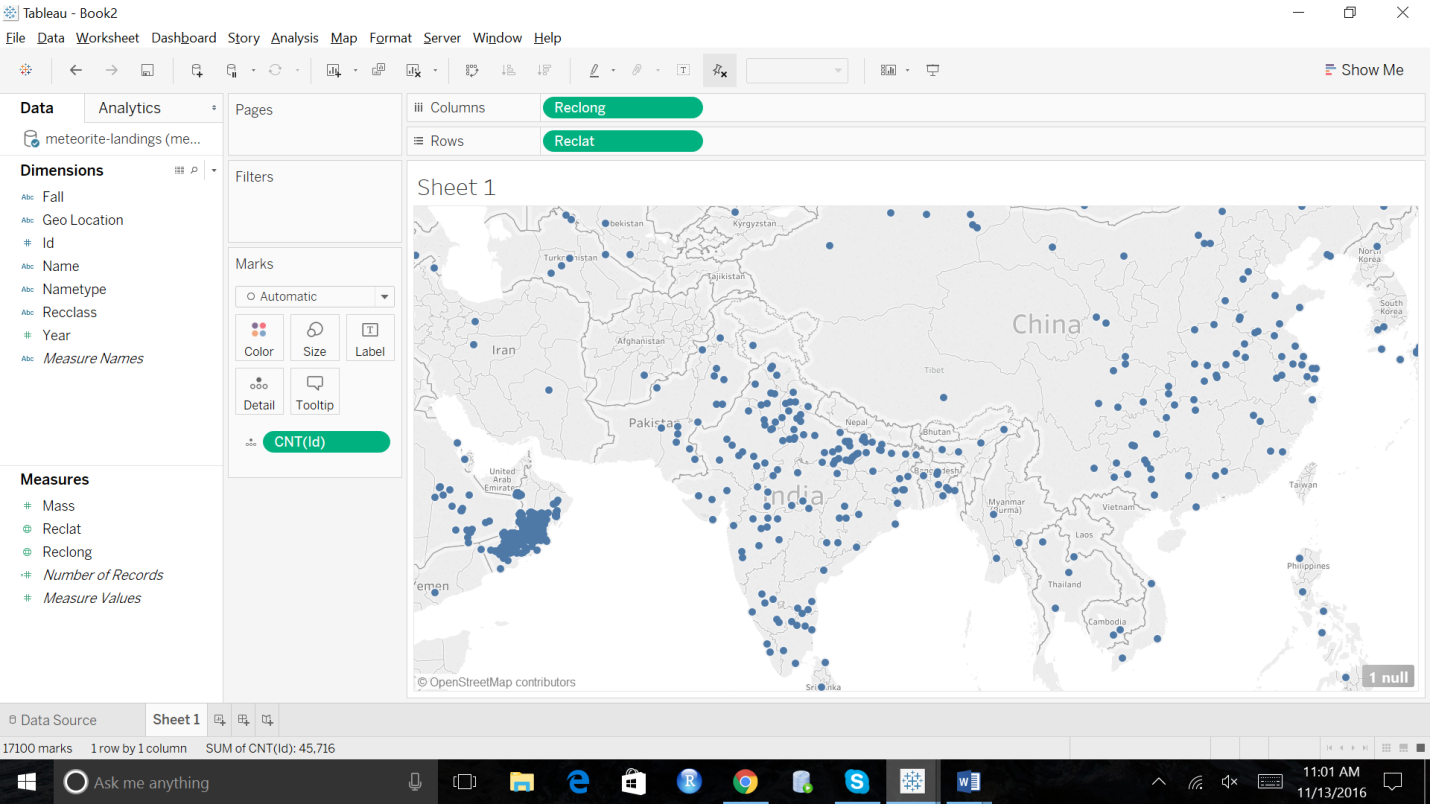
As per the R code which was given by a person who as analyzed this data on Kaggle, there are no meteorites which were either found or fell in India. However, when we analyzed the data we noticed that the locations specified may not be correct. Hence, we loaded the data in TABLEAU and noticed that there are considerably a large number to meteorites present in India as well.

Also, as per his analysis on Kaggle, there are more than 60% meteorites, which belong to Antarctica region only, however, we contradict this and as per our analysis we do not see any meteorites at all.

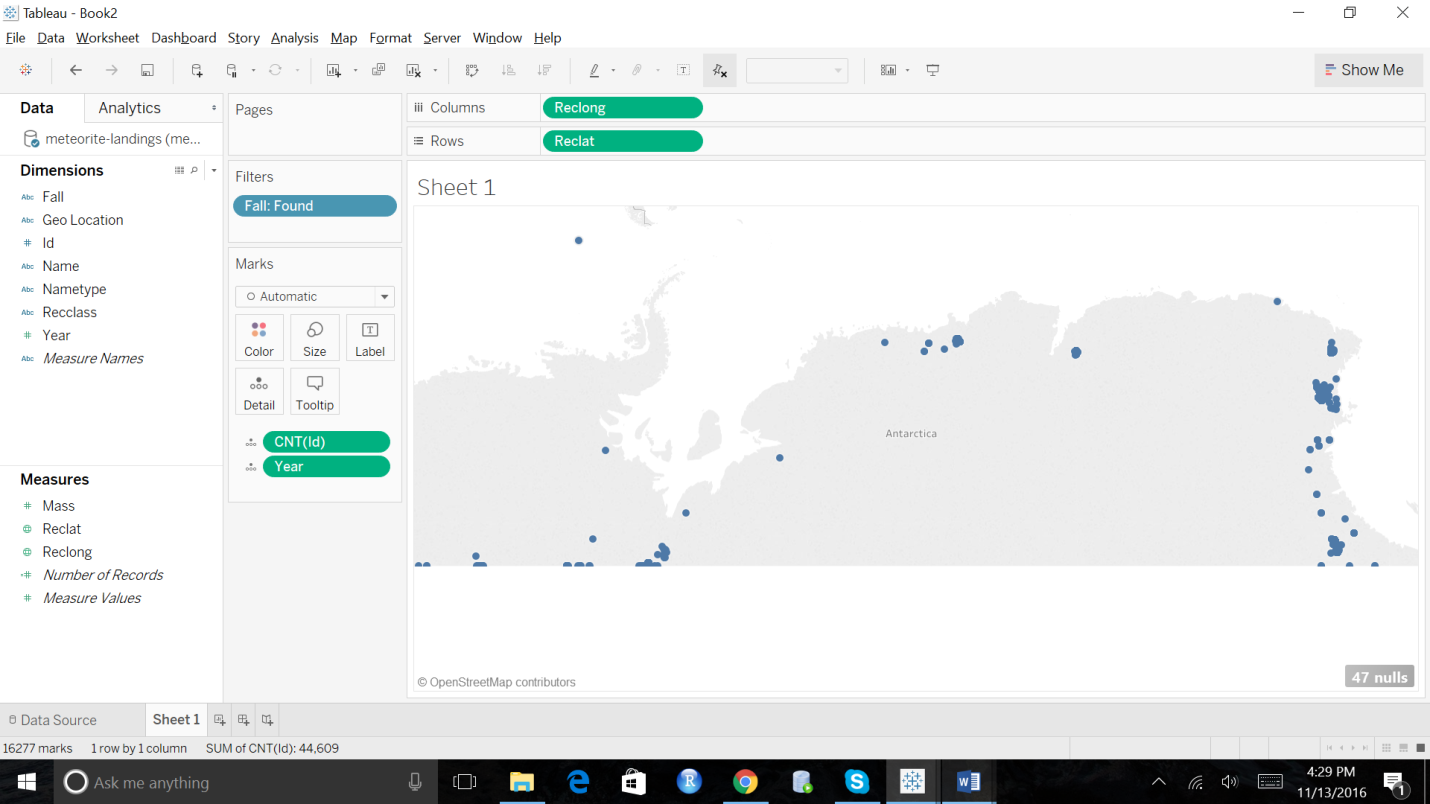
We observed that he as compared the data of longitude and latitude with the world data. This data may not be accurate or old.

Below are the screenshots which describes the same -

1. Below image showing that there are meteorites in India as well.



1. No Meteorites in Antarctica –

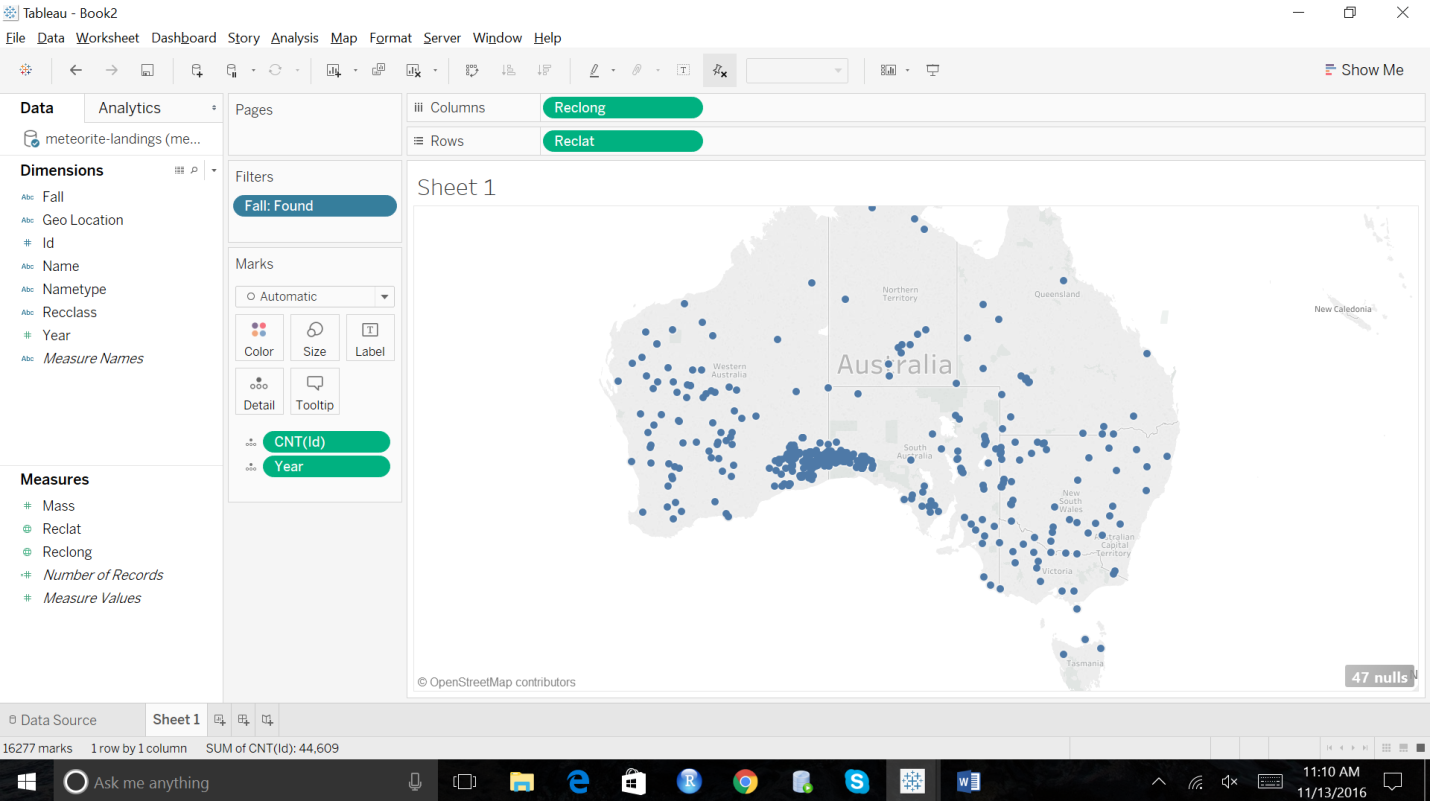


1. Analysis on Australia –

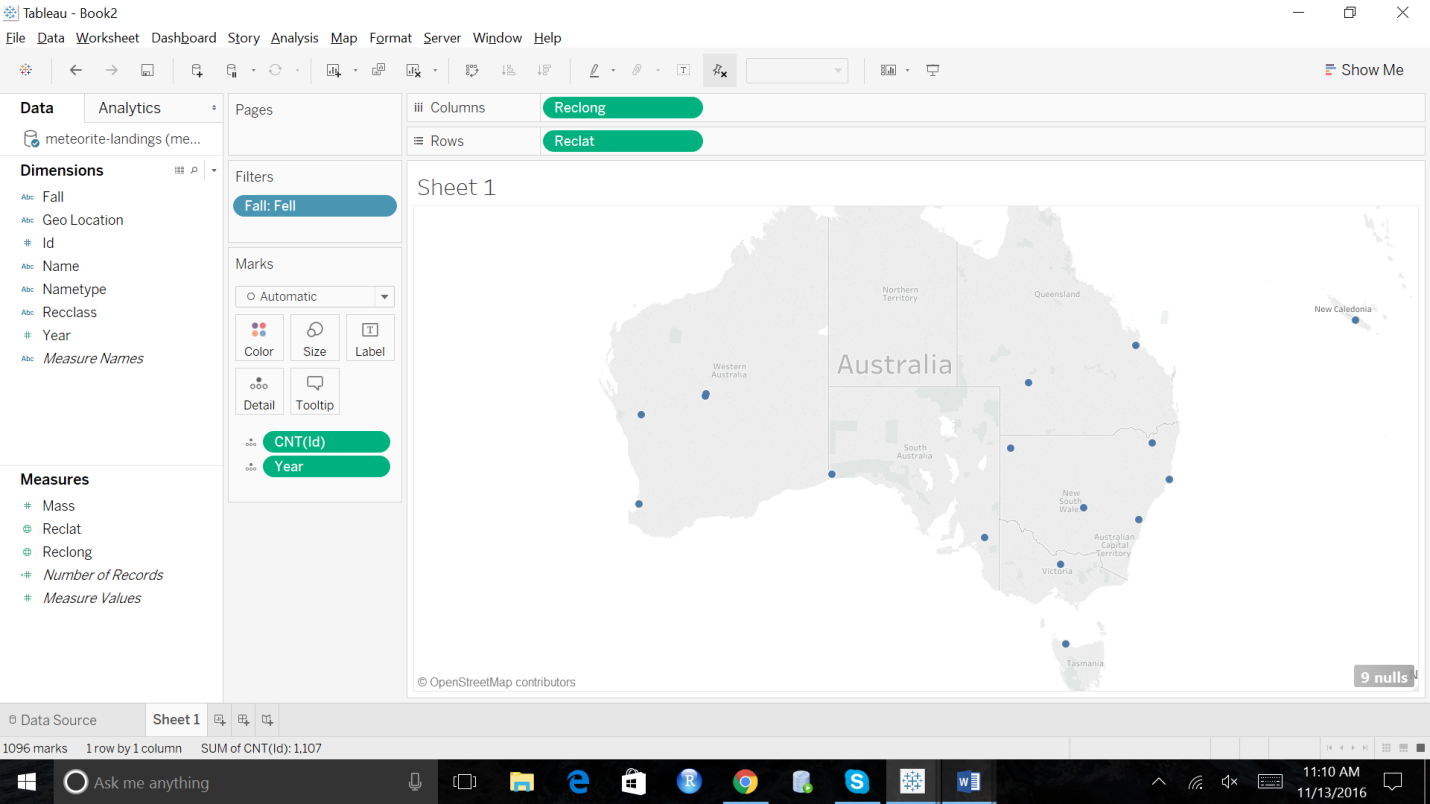
In the procedure of finding a unique country to perform analysis, we have noticed that there was a well-defined pattern of these meteorites. However, we have noticed that the majority of these were found and a very few were fall.

As per our analysis on the data, we preferred not to consider the found as this means that the meteorite was found after it fell, may be after few years, and fall means that we have noticed the meteorite while it was falling.

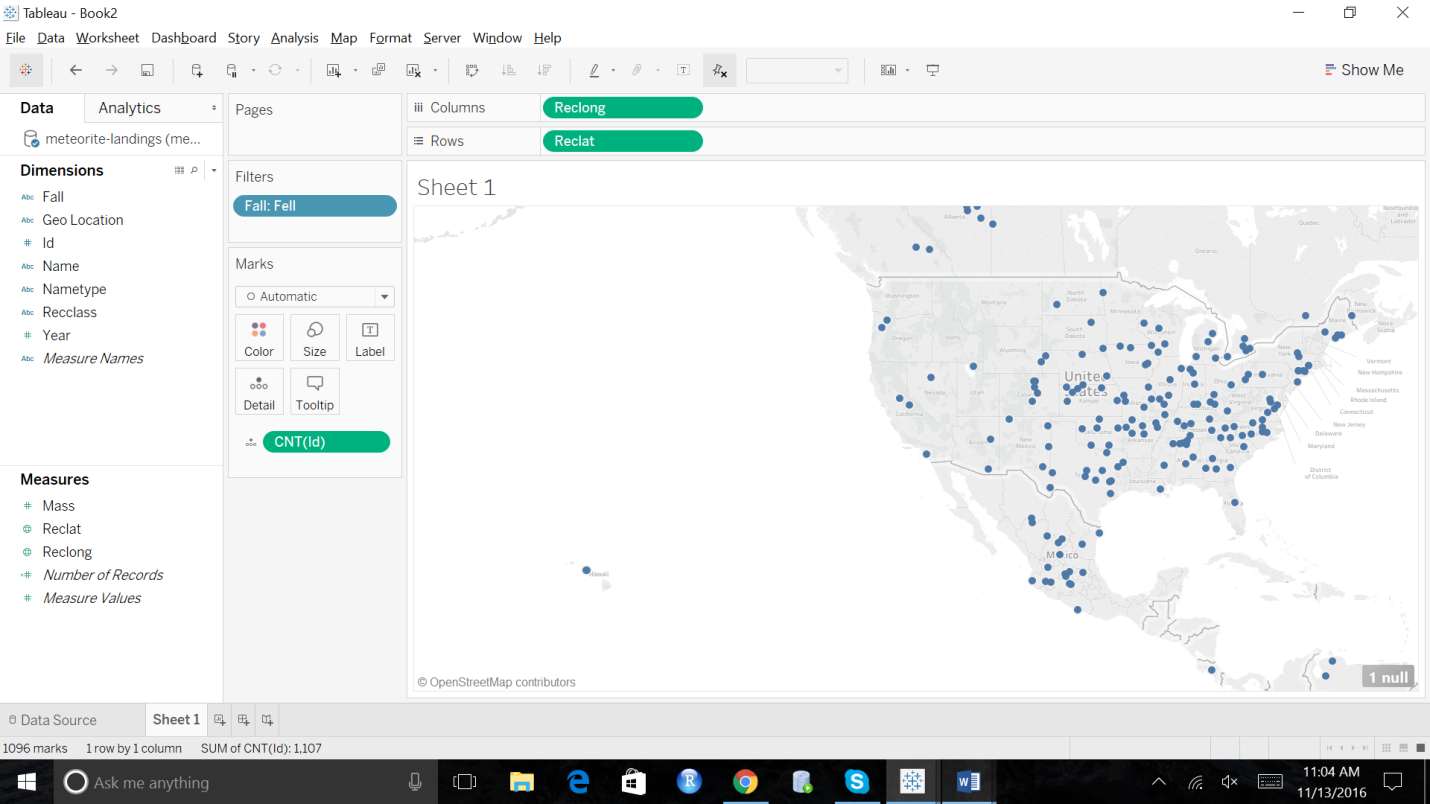
Screenshot 3 – Number of meteorites of Found type.



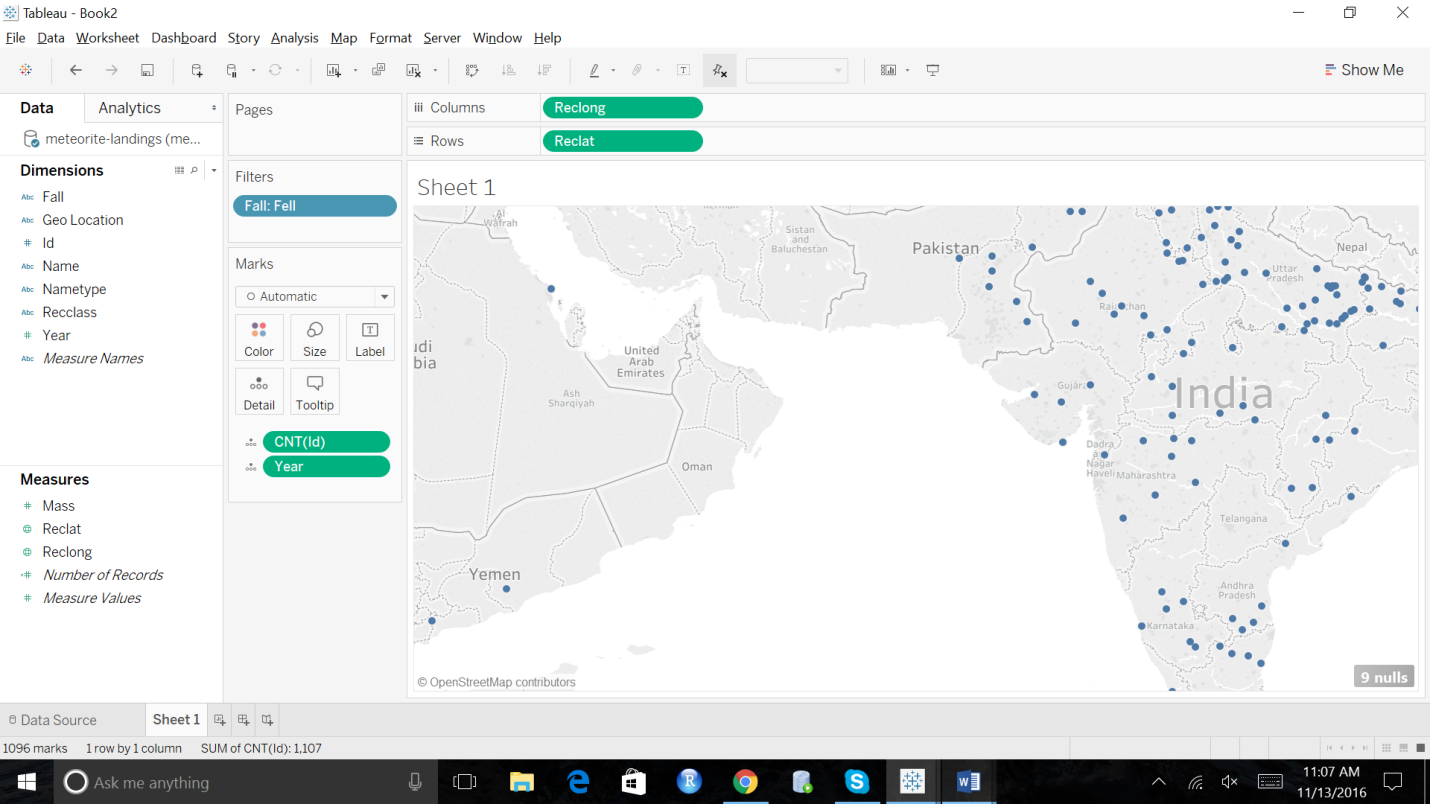
Type - Fell

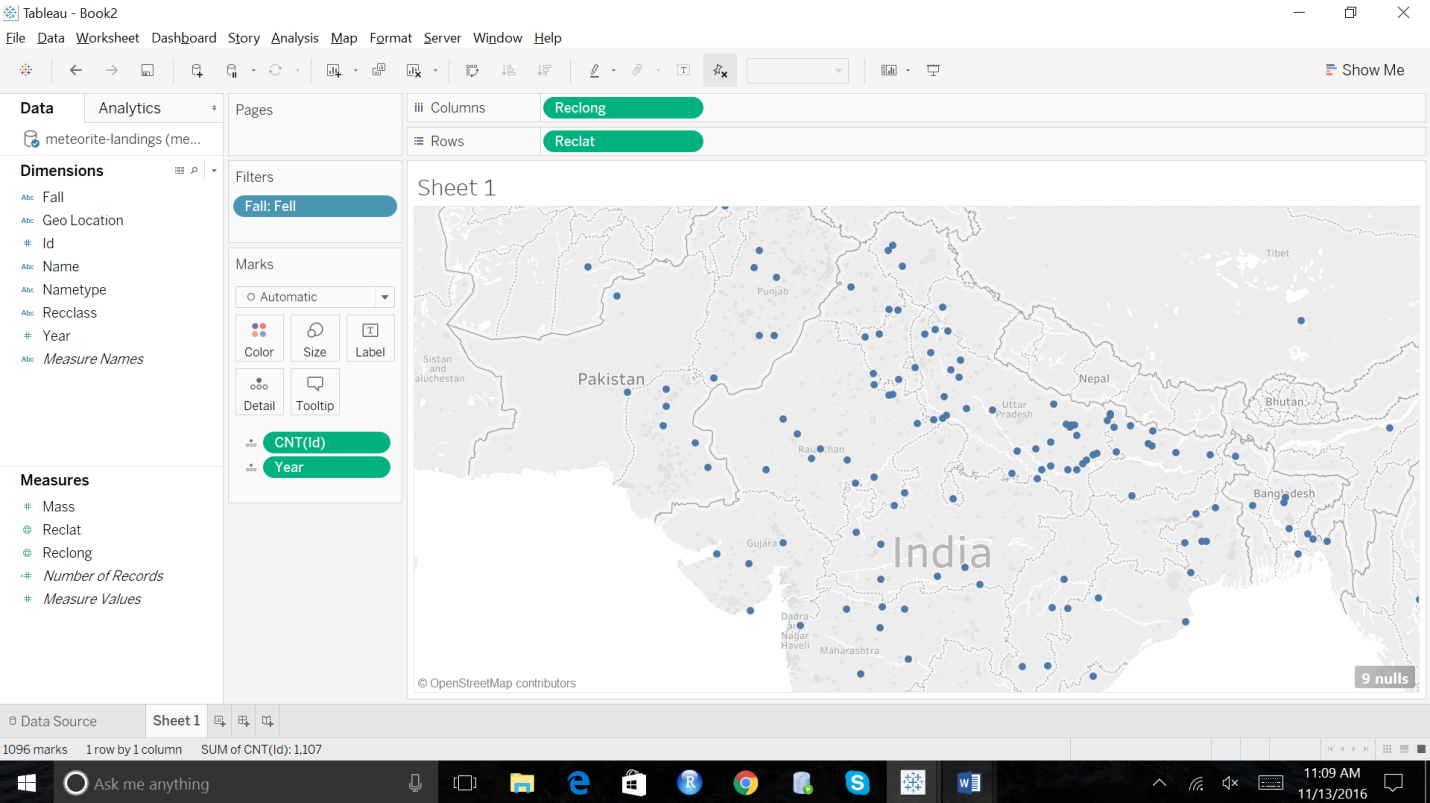


Screenshot#5 - of Fells in USA.

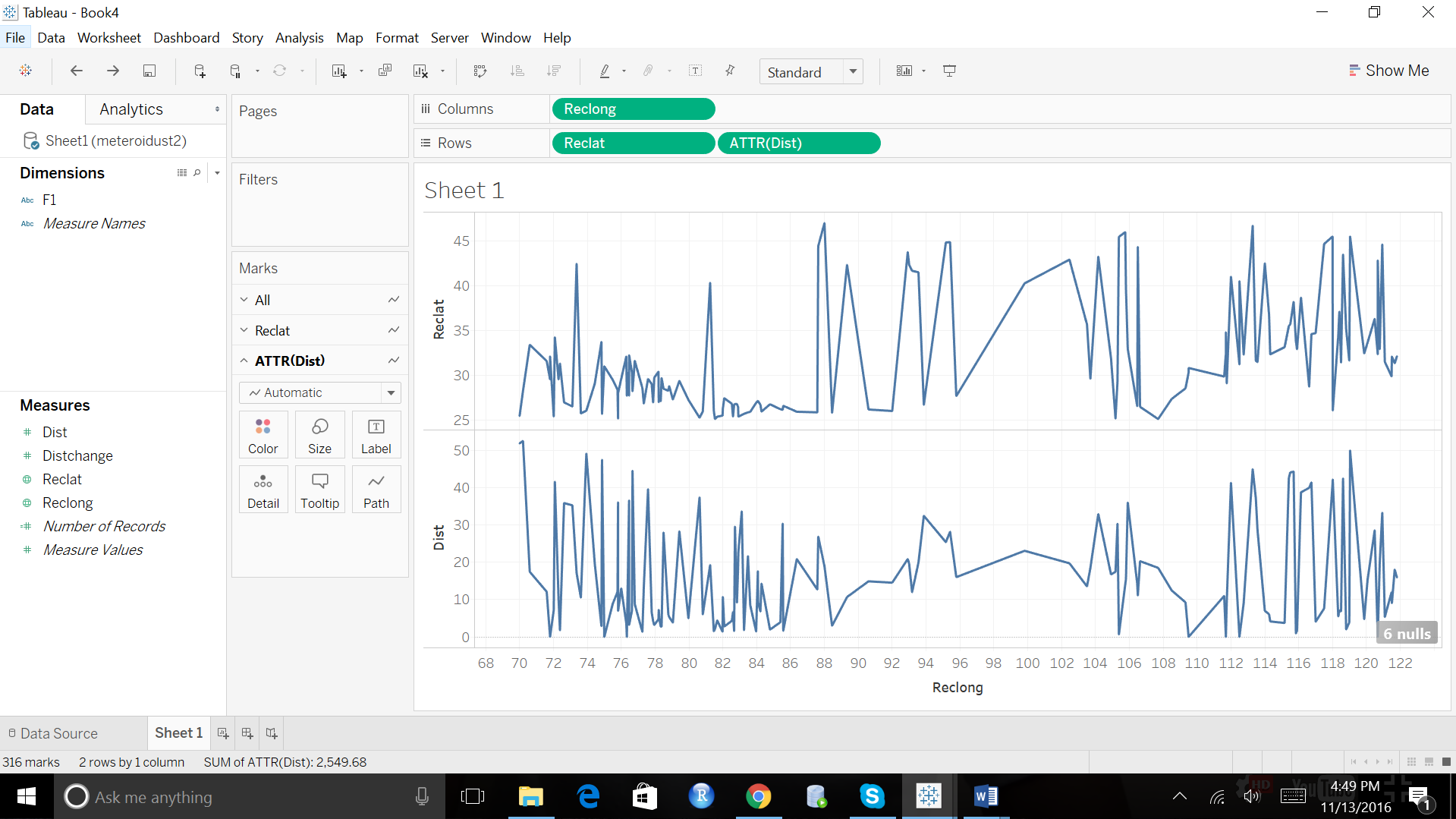


Screenshot#6 – We have considered one country exclusively for doing our analysis, India. Below screenshots show the distribution of meteorites in India.

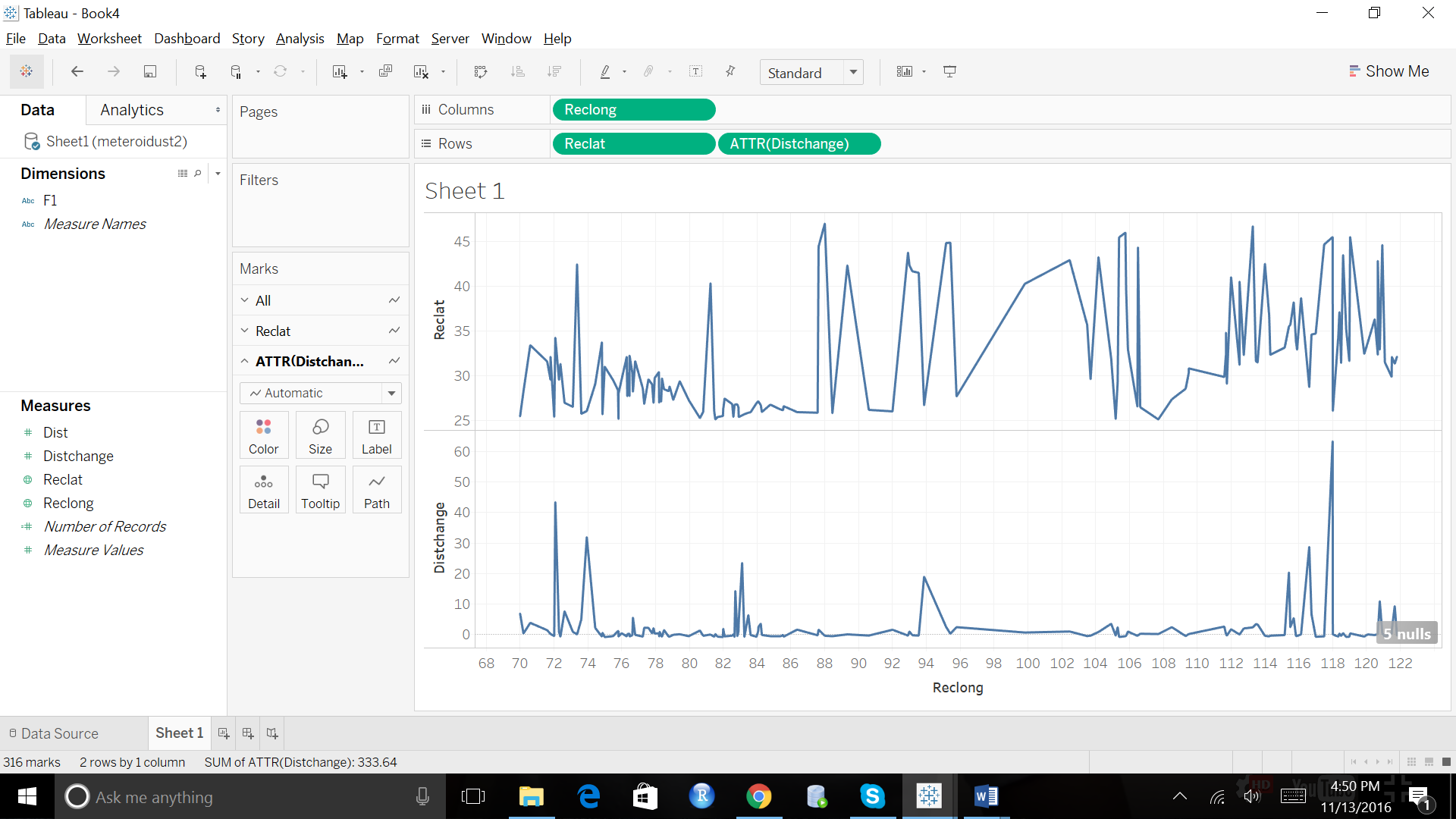




Below screenshot shows the change of distance with respect to year for India.



Shows the distance change –



**Conclusion:**

1. Through this, we have found out the average mass of the meteorite that would hit the earth for next years.
2. We have also found out the Average distance of location that would be a probable location for the next meteorite fall for the geo locations of INDIA.
3. We have found the pattern of the location of geo codes and their movement by providing the average of the change in distance between the consecutive locations.