quality check of meta location table for JHU COVID19 data

```
# check the consistency of worldmet station coordiation and JHU coordinates
rm(list=ls())
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.1
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
metaTb = read_csv("_metaTb_1.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   UID = col_double(),
##   iso2 = col_character(),
##   iso3 = col_character(),
##   code3 = col_double(),
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Combined_Key = col_character(),
##   latStation = col_double(),
##   longStation = col_double(),
##   usaf = col_character(),
##   wban = col_character(),
##   station = col_character()
## )
```

```
head(metaTb)
```

```
## # A tibble: 6 x 17
##      X1    UID iso2  iso3  code3  FIPS Admin2 Province_State Country_Region
##   <dbl>  <dbl> <chr> <chr> <dbl> <dbl> <chr>  <chr>          <chr>
## 1     1 8.40e7 US    USA     840  1001 Autau~ Alabama        US
## 2     2 8.40e7 US    USA     840  1003 Baldw~ Alabama        US
## 3     3 8.40e7 US    USA     840  1005 Barbo~ Alabama        US
## 4     4 8.40e7 US    USA     840  1007 Bibb   Alabama        US
## 5     5 8.40e7 US    USA     840  1009 Blount Alabama        US
```

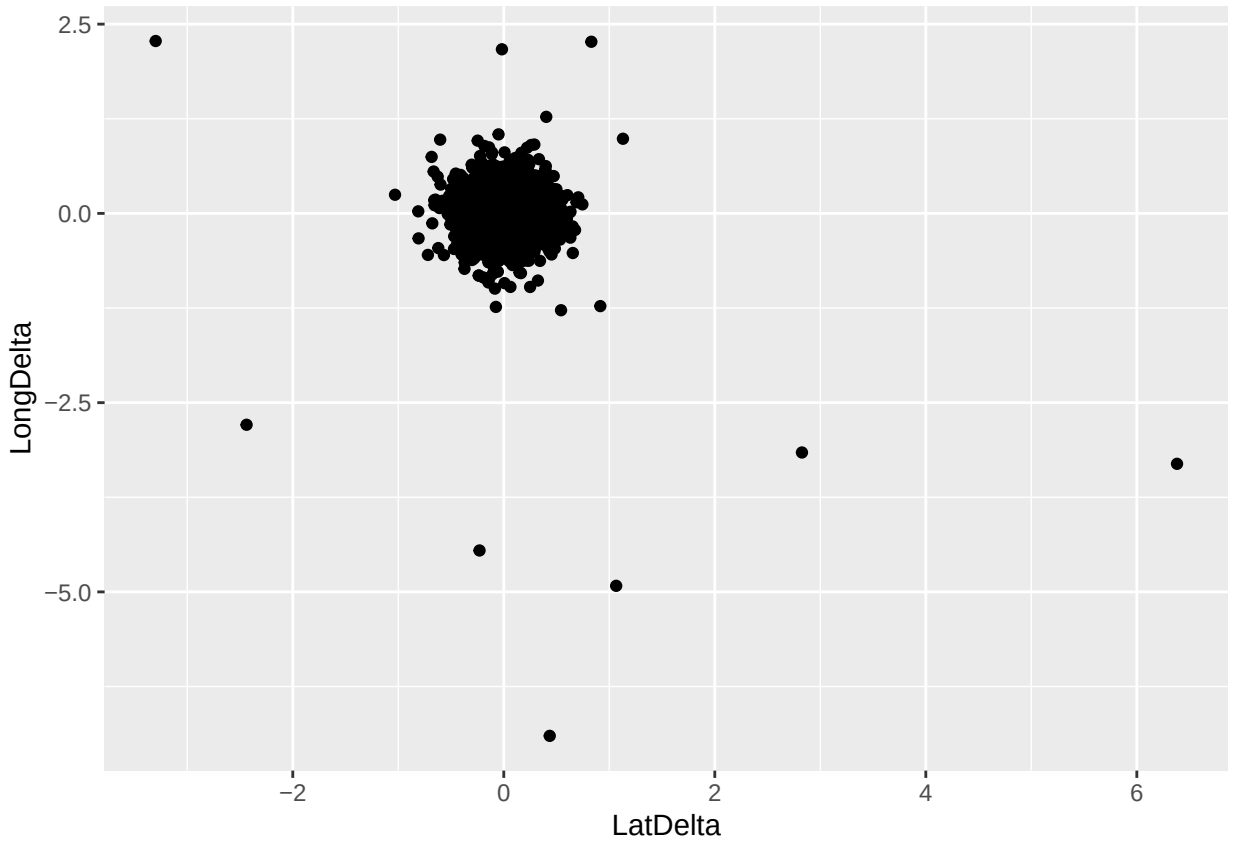
```
## 6     6 8.40e7 US    USA     840  1011 Bullo~ Alabama        US
## # ... with 8 more variables: Lat <dbl>, Long_ <dbl>, Combined_Key <chr>,
## #   latStation <dbl>, longStation <dbl>, usaf <chr>, wban <chr>, station <chr>
```

```
#summary(metaTb$FIPS)
# there are places without FIPS
metaTb[which(is.na(metaTb$FIPS)), ]
```

```
## # A tibble: 10 x 17
##       X1    UID iso2  iso3  code3  FIPS Admin2 Province_State Country_Region
##    <dbl>  <dbl> <chr> <chr> <dbl> <dbl> <chr>  <chr>          <chr>
##  1  1268 8.41e7 US    USA     840    NA Dukes~ Massachusetts  US
##  2  1305 8.41e7 US    USA     840    NA Feder~ Michigan       US
##  3  1337 8.41e7 US    USA     840    NA Michi~ Michigan       US
##  4  1592 8.41e7 US    USA     840    NA Kansa~ Missouri       US
##  5  2955 8.41e7 US    USA     840    NA Bear ~ Utah           US
##  6  2960 8.41e7 US    USA     840    NA Centr~ Utah           US
##  7  2979 8.41e7 US    USA     840    NA South~ Utah           US
##  8  2980 8.41e7 US    USA     840    NA South~ Utah           US
##  9  2983 8.41e7 US    USA     840    NA TriCo~ Utah           US
## 10  2991 8.41e7 US    USA     840    NA Weber~ Utah           US
## # ... with 8 more variables: Lat <dbl>, Long_ <dbl>, Combined_Key <chr>,
## #   latStation <dbl>, longStation <dbl>, usaf <chr>, wban <chr>, station <chr>
```

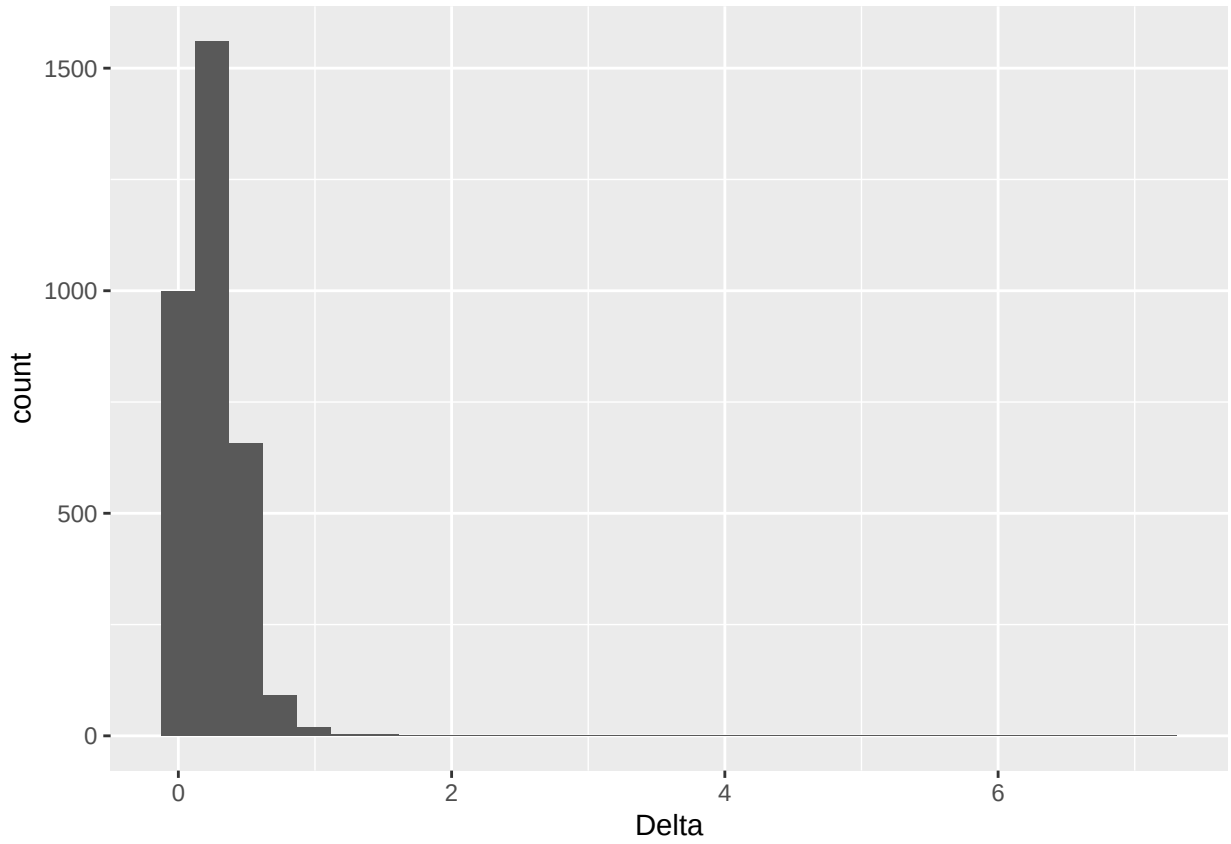
```
metaTb$LatDelta =  metaTb$Lat - metaTb$latStation
metaTb$LongDelta = metaTb$Long_ - metaTb$longStation
metaTb$Delta = sqrt( metaTb$LatDelta* metaTb$LatDelta + metaTb$LongDelta * metaTb$LongDelta )

ggplot( metaTb, aes(x=LatDelta, y=LongDelta)) + geom_point()
```




```
ggplot(data=metaTb, aes(x=Delta)) +   geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# 1 degree latitue is about 69 miles, 111 kilometers
# it seems 0.5 degree is a good cutoff, about 90% quantile

quantile( metaTb$Delta, prob=c(0.05, 0.9, 0.95, 0.97, 0.99))
```

```
##         5%        90%        95%        97%        99%
## 0.02704341 0.49549495 0.58310290 0.65809015 0.86452903
```

```
Deltacutoff = 0.5

metaTb$StationProximity = ifelse( metaTb$Delta >= Deltacutoff, FALSE, TRUE )
summary(metaTb$StationProximity)
```

```
##    Mode   FALSE    TRUE
## logical     325    3015
```

```
summary(metaTb)
```

```
##        X1              UID               iso2              iso3
##  Min.   :   1.0   Min.   :      16   Length:3340        Length:3340
##  1st Qu.: 835.8   1st Qu.:84018108   Class :character   Class :character
##  Median :1670.5   Median :84029208   Mode  :character   Mode  :character
##  Mean   :1670.5   Mean   :83429580
##  3rd Qu.:2505.2   3rd Qu.:84046120
##  Max.   :3340.0   Max.   :84099999
##
##      code3            FIPS          Admin2          Province_State
##  Min.   : 16.0   Min.   :   60   Length:3340        Length:3340
##  1st Qu.:840.0   1st Qu.:19080   Class :character   Class :character
```

```
##  Median :840.0   Median :31014   Mode  :character   Mode  :character
##  Mean   :834.5   Mean   :33062
##  3rd Qu.:840.0   3rd Qu.:47130
##  Max.   :850.0   Max.   :99999
##                  NA's   :10
##  Country_Region          Lat             Long_          Combined_Key
##  Length:3340        Min.   :-14.27   Min.   :-174.16   Length:3340
##  Class :character   1st Qu.: 33.90   1st Qu.: -97.79   Class :character
##  Mode  :character   Median : 38.00   Median : -89.49   Mode  :character
##                     Mean   : 36.71   Mean   : -88.60
##                     3rd Qu.: 41.57   3rd Qu.: -82.31
##                     Max.   : 69.31   Max.   : 145.67
##
##    latStation      longStation          usaf               wban
##  Min.   :-14.28   Min.   :-174.17   Length:3340        Length:3340
##  1st Qu.: 33.91   1st Qu.: -97.73   Class :character   Class :character
##  Median : 38.00   Median : -89.41   Mode  :character   Mode  :character
##  Mean   : 36.70   Mean   : -88.59
##  3rd Qu.: 41.57   3rd Qu.: -82.37
##  Max.   : 68.48   Max.   : 145.72
##
##    station             LatDelta           LongDelta             Delta
##  Length:3340        Min.   :-3.299209   Min.   :-6.901629   Min.   :0.0000
##  Class :character   1st Qu.:-0.094563   1st Qu.:-0.126027   1st Qu.:0.1053
##  Mode  :character   Median : 0.000000   Median : 0.000000   Median :0.2239
##                     Mean   : 0.003277   Mean   :-0.007263   Mean   :0.2631
##                     3rd Qu.: 0.101684   3rd Qu.: 0.121500   3rd Qu.:0.3620
##                     Max.   : 6.381041   Max.   : 2.277272   Max.   :7.1876
##
##  StationProximity
##  Mode :logical
##  FALSE:325
##  TRUE :3015
##
##
##
##
```

```
metaTb$GoodLocation = metaTb$StationProximity
metaTb$GoodLocation = ifelse( is.na(metaTb$FIPS) | metaTb$FIPS==999999 | metaTb$station=='BOGUS CHINESE
```

```
metaTb[171,]
```

```
## # A tibble: 1 x 22
##      X1    UID iso2  iso3  code3  FIPS Admin2 Province_State Country_Region
##   <dbl>  <dbl> <chr> <chr> <dbl> <dbl> <chr>  <chr>          <chr>
## 1   171 8.41e7 US    USA     840 80005 Out o~ Arkansas       US
## # ... with 13 more variables: Lat <dbl>, Long_ <dbl>, Combined_Key <chr>,
## #   latStation <dbl>, longStation <dbl>, usaf <chr>, wban <chr>, station <chr>,
## #   LatDelta <dbl>, LongDelta <dbl>, Delta <dbl>, StationProximity <lgl>,
## #   GoodLocation <lgl>
```

```
write_csv(metaTb, "_metaStation_20201027b.csv")
```