Selecting lines for HiLo F1s

Daniel Runcie

Start with SeeD GWAS GBS lines

I want all lines we select to be here to be able to take advantage of potential genotype info

Load the SeeDs data and match up with the CIMMYT data to extract CIMMYT availability and IDs

Load CIMMYT data for availability:

CIMMYT data and name conversion table, match up, and add bank numbers to CIMMYT_data table

Add CIMMYT info to SeeD:

Load GRIN data for Lat/Long/Elevation of some lines

Load GRIN data

Identify lines with multiple CIMMYTMA IDs. Exclude these lines. Numbers of lines with 0-2 CIMMYTMA IDs:

```
##     0     1     2
## 36021  9878    28
```

Identify lines that share the same CIMMYTMA ID. Exclude these lines.

```
## [1] 121
```

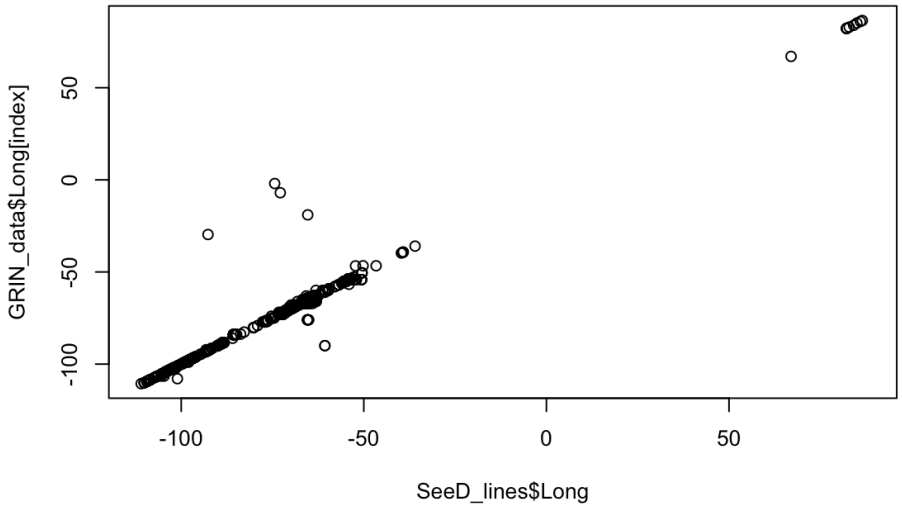
Match up GRIN and SeeD and add missing Elevation.

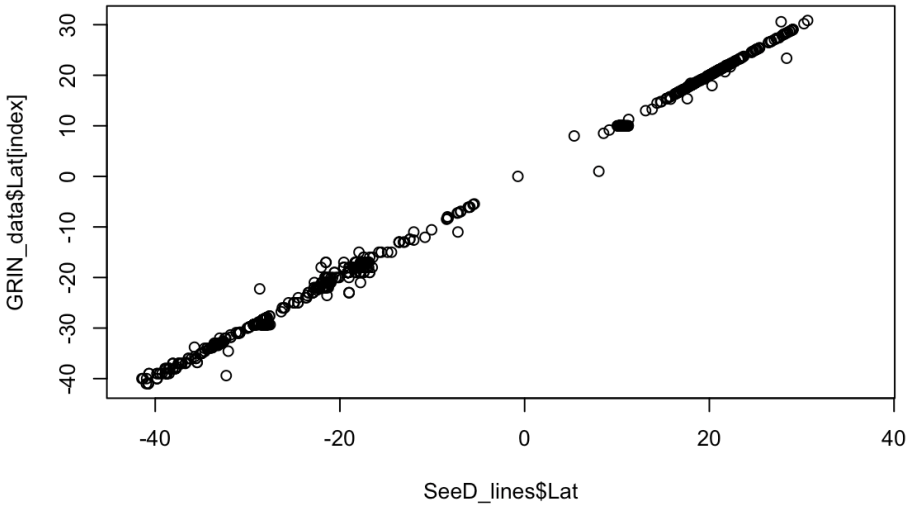
When info differs for the same line, use SeeD for Lat/Long, and GRIN for elevation

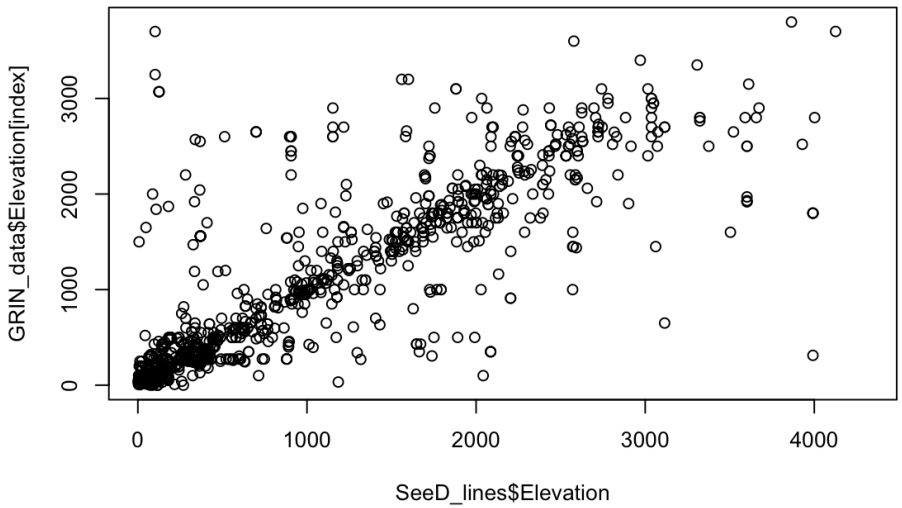
```
## [1] 0
```

```
## [1] 0
```

```
## [1] 44
```


GRIN_data$Long[index]
-100
-50
0
50
-100
-50
0
50
SeeD_lines$Long


GRIN_data$Lat[index]
-40
-30
-20
-10
0
10
20
30
-40
-20
0
20
40
SeeD_lines$Lat


GRIN_data$Elevation[index]
0
1000
2000
3000
0
1000
2000
3000
4000
SeeD_lines$Elevation

Drop lines without coordinates

Load Current F1s

Make all available lines database

```
## [1] 16120     7
```

Which lines have GBS?

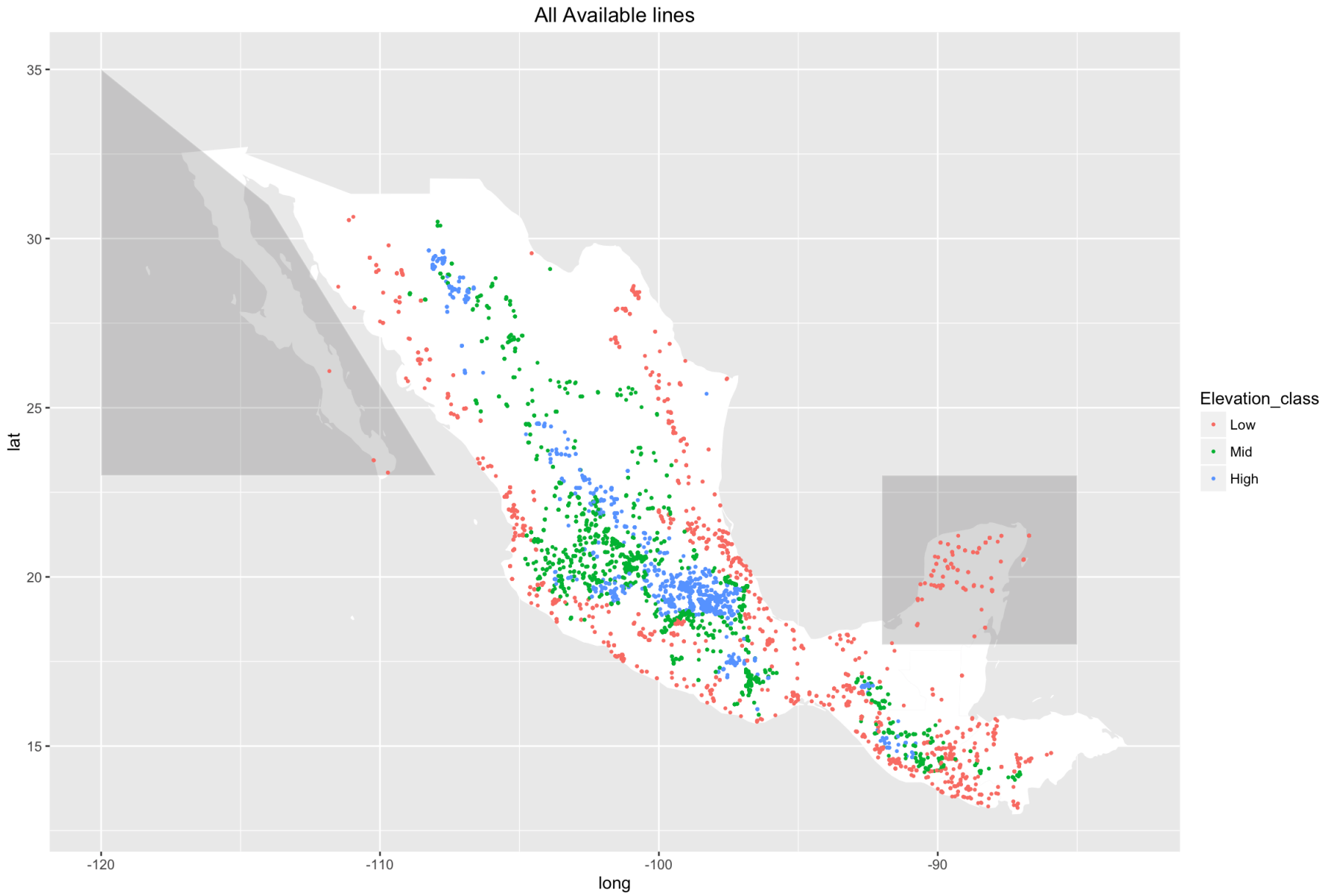
Divide lines by elevation:

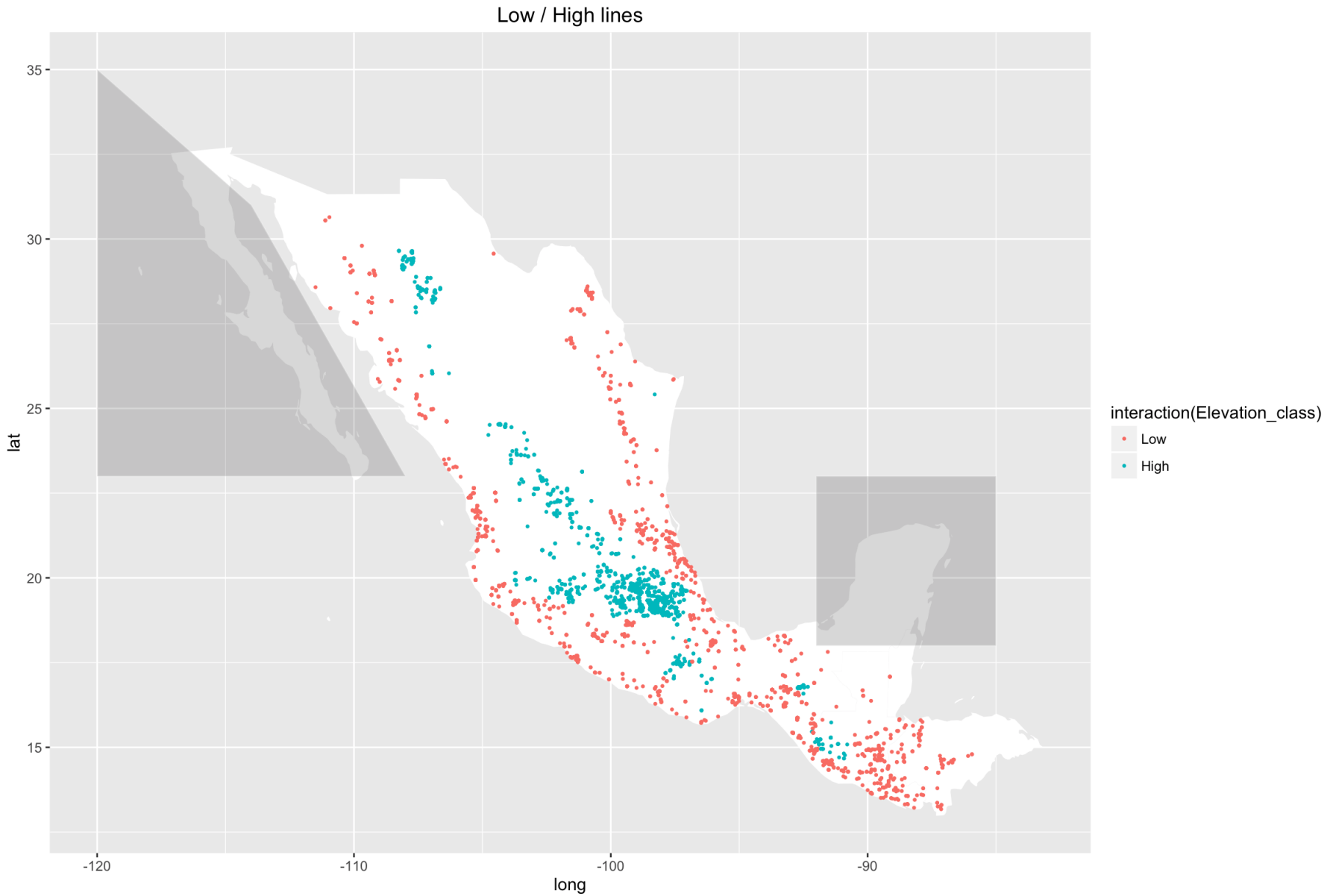
Subset lines by regions (Central America vs South America). Plot

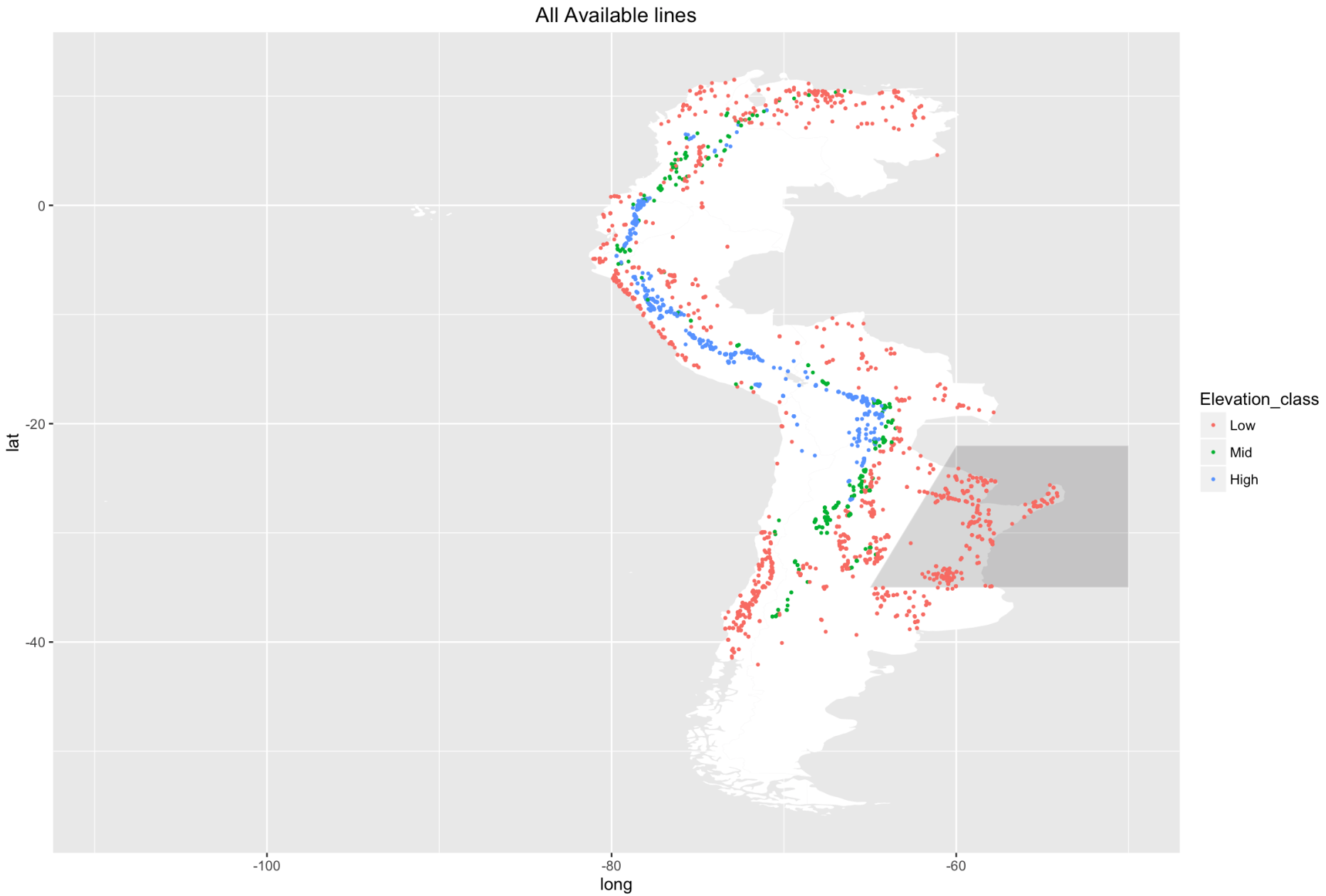
Selecting candidate lines

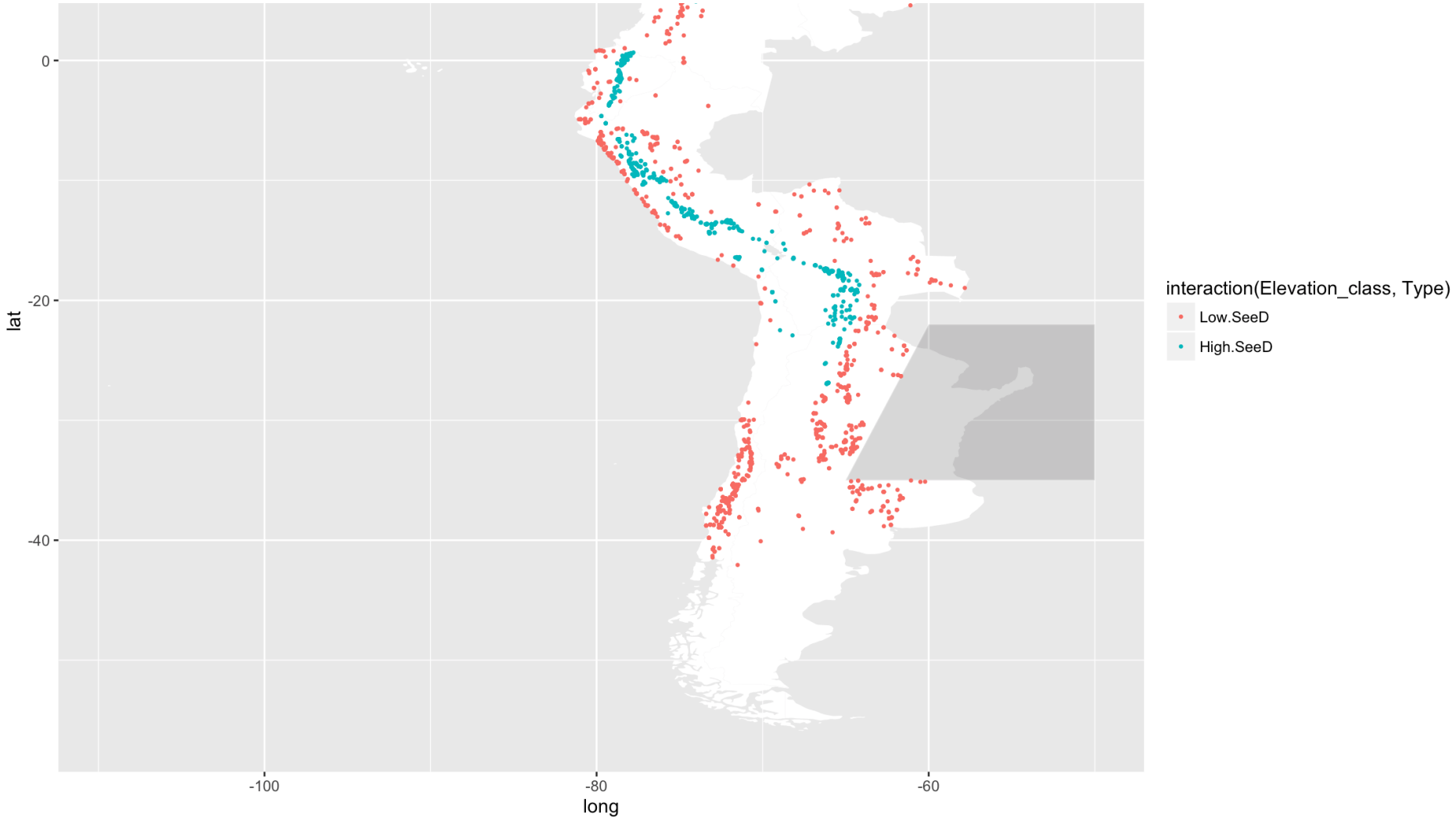
Selection based on:

1. Available in CIMMYT
2. Either < 1000m or > 2000m
3. In SeeD database (although elevation info may have come from GRIN)


All Available lines
35
30
25
20
15
lat
-120
-110
-100
-90
long
Elevation_class
Low
Mid
High


Low / High lines
35
30
25
20
15
lat
-120
-110
-100
-90
long
interaction(Elevation_class)
Low
High


All Available lines
0
-20
-40
lat
-100
-80
-60
long
Elevation_class
Low
Mid
High
Low / High lines



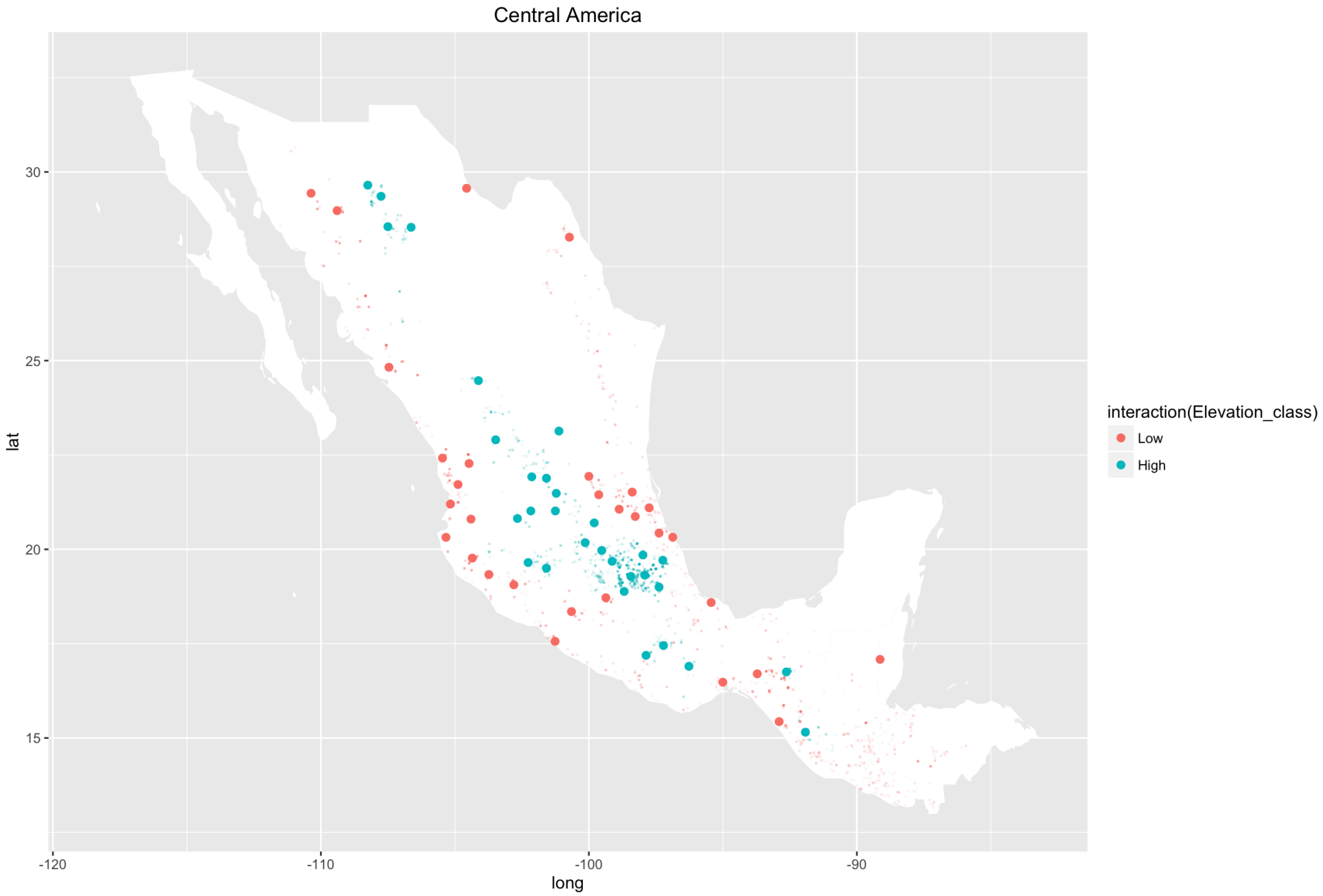
Selection algorithm 1

In this algorithm, I randomly select a number of high elevation lines, with the restriction that they cannot be too close together. Then for each high line, I pick a low line in the same latitude band, also not too close to any other line

Select Mex lines

```
##       SeeD Current_F1
##         56          4
```

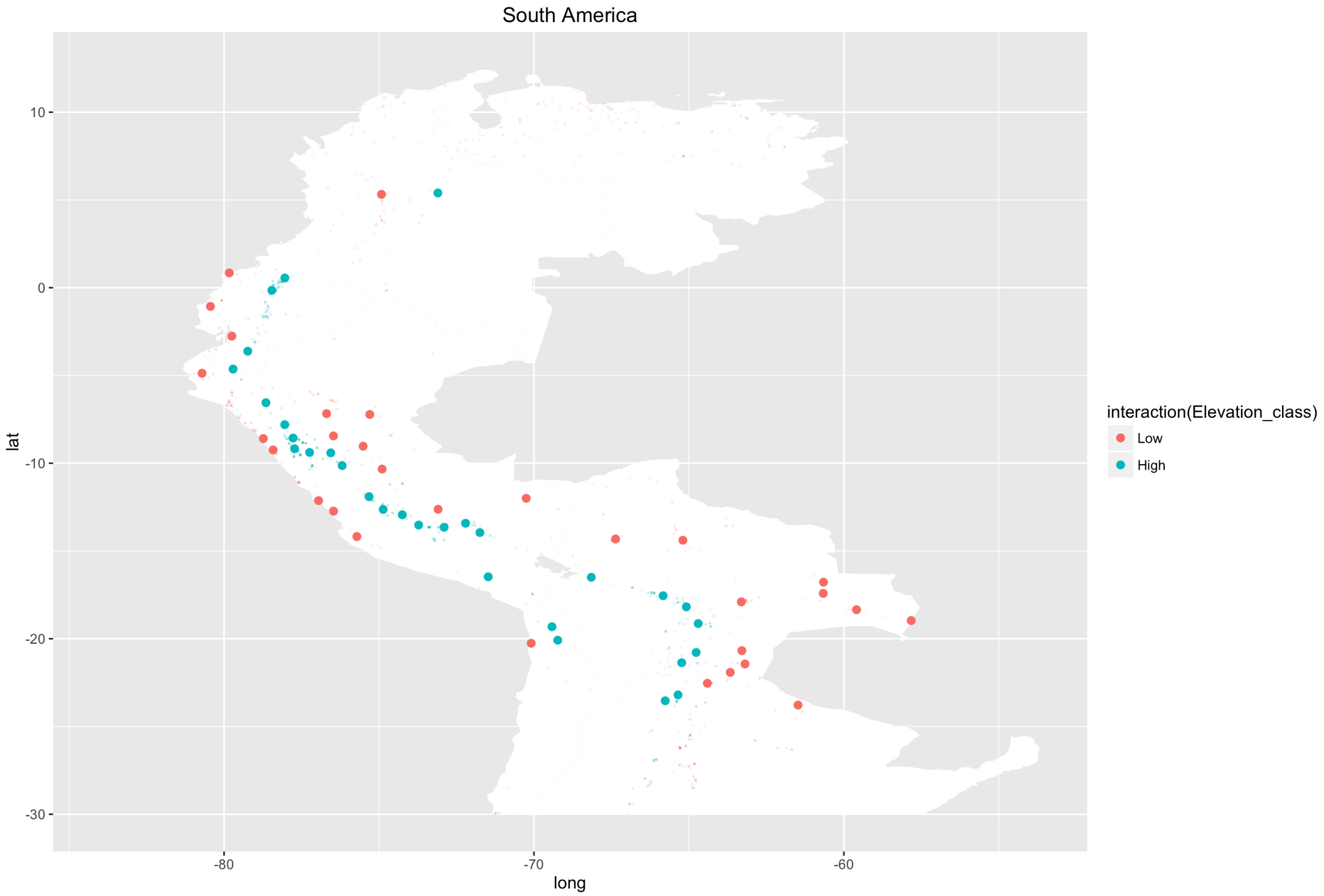
```
##  Low  Mid High
##   30    0   30
```



Select SA lines

```
##       SeeD Current_F1
##         60           0
```

```
##  Low  Mid High
##   30    0   30
```



Save lines in a DF

```
##                ID           Pair          Long              Lat
##  CIMMYTMA-000449:  1   Mex_1  :  2   Min.   :-110.37   Min.   :-23.780
##  CIMMYTMA-000547:  1   Mex_10 :  2   1st Qu.:-100.67   1st Qu.:-12.783
##  CIMMYTMA-000554:  1   Mex_11 :  2   Median : -84.92   Median : 10.276
##  CIMMYTMA-000645:  1   Mex_12 :  2   Mean   : -86.24   Mean   :  4.357
##  CIMMYTMA-000686:  1   Mex_13 :  2   3rd Qu.: -73.57   3rd Qu.: 20.498
##  CIMMYTMA-000861:  1   Mex_14 :  2   Max.   : -57.83   Max.   : 29.650
##  (Other)        :114   (Other):108
##    Elevation           Race       GBS          Elevation_class
##  Min.   :   2.0          :45   Mode :logical   Low :60
##  1st Qu.: 303.8   TUXPEN :10   FALSE:107       Mid : 0
##  Median :1505.5   CELAYA : 6   TRUE :13        High:60
##  Mean   :1443.4   CUBAAM : 6   NA's :0
##  3rd Qu.:2404.5   CONICO : 5
##  Max.   :3929.0   CONNOR : 5
##                   (Other):43
##        id        general_identifier              bank_number
##  Min.   :  189   Min.   :  5474     CIMMYTMA-BANK-000449:  1
##  1st Qu.:13690   1st Qu.:248942     CIMMYTMA-BANK-000547:  1
##  Median :18102   Median :253356     CIMMYTMA-BANK-000554:  1
##  Mean   :17205   Mean   :224272     CIMMYTMA-BANK-000645:  1
##  3rd Qu.:22148   3rd Qu.:257401     CIMMYTMA-BANK-000686:  1
##  Max.   :30844   Max.   :267450     (Other)             :111
##  NA's   :4       NA's   :4          NA's                :  4
##      collnumb         colldate
##  AGUC67  :  1            :15
##  ANC.526 :  1   1954-01-01: 8
##  APUC.169:  1   1968-01-01: 5
##  AREQ.164:  1   1943-01-01: 3
##  ARGE521 :  1   1947-01-01: 3
##  (Other) :111   (Other)   :82
##  NA's    :  4   NA's      : 4
##  Sample.ID.of.DNA.from.single.plants.used.in.GWAS          CIMMYT_ID
##              :103                                  CIMMYTMA 10248:  1
##  SEEDGWAS1031:  1                                  CIMMYTMA 11854:  1
##  SEEDGWAS156 :  1                                  CIMMYTMA 11857:  1
##  SEEDGWAS2628:  1                                  CIMMYTMA 12085:  1
##  SEEDGWAS2678:  1                                  CIMMYTMA 12231:  1
```

```
##  (Other)     :  9                                  CIMMYTMA 12290:  1
##  NA's        :  4                                  (Other)       :114
##  SeeD_Elevation   Elevation_flag
##  Min.   :   2.0   Mode :logical
##  1st Qu.: 319.5   FALSE:116
##  Median : 973.0   NA's :4
##  Mean   :1428.4
##  3rd Qu.:2421.0
##  Max.   :3929.0
##  NA's   :4
```