

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
```

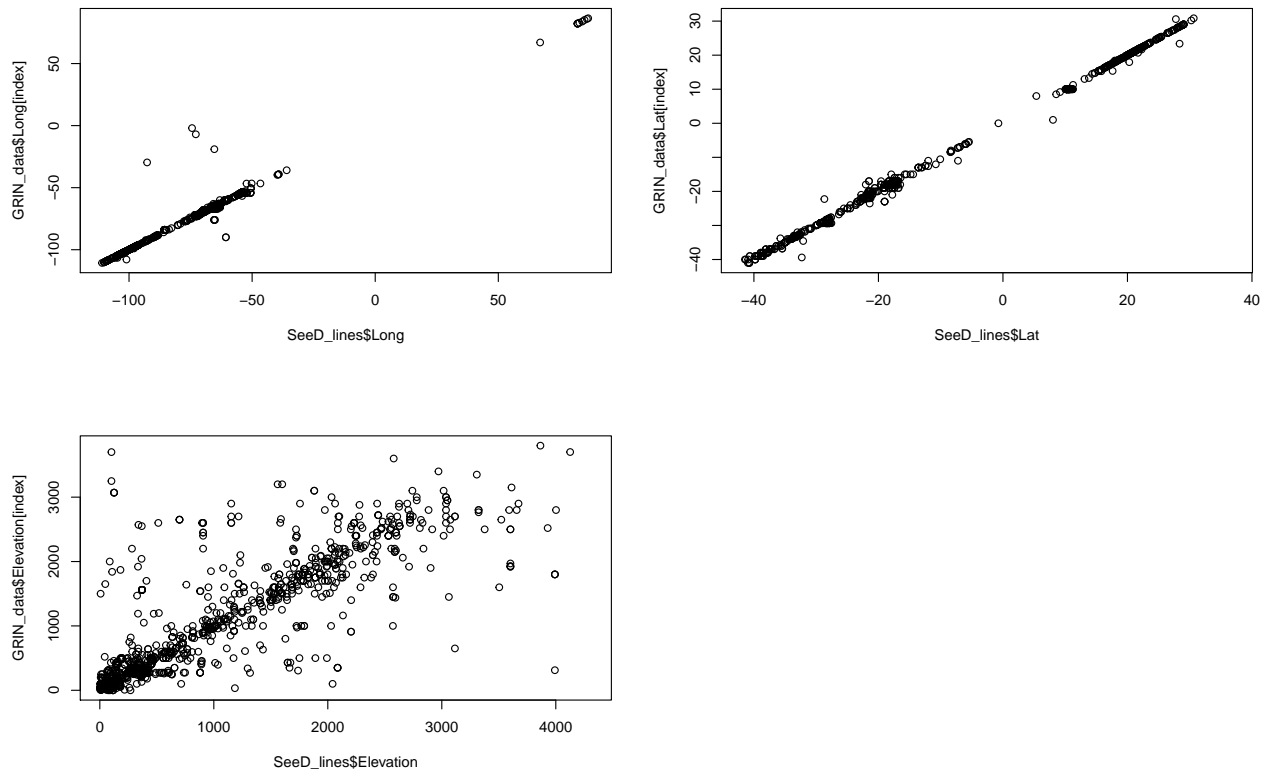
Match up GRIN and SeeD and add missing Elevation.

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

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

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

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



Drop lines without coordinates

Load Current F1s

Make all available lines database

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

Which lines have GBS?

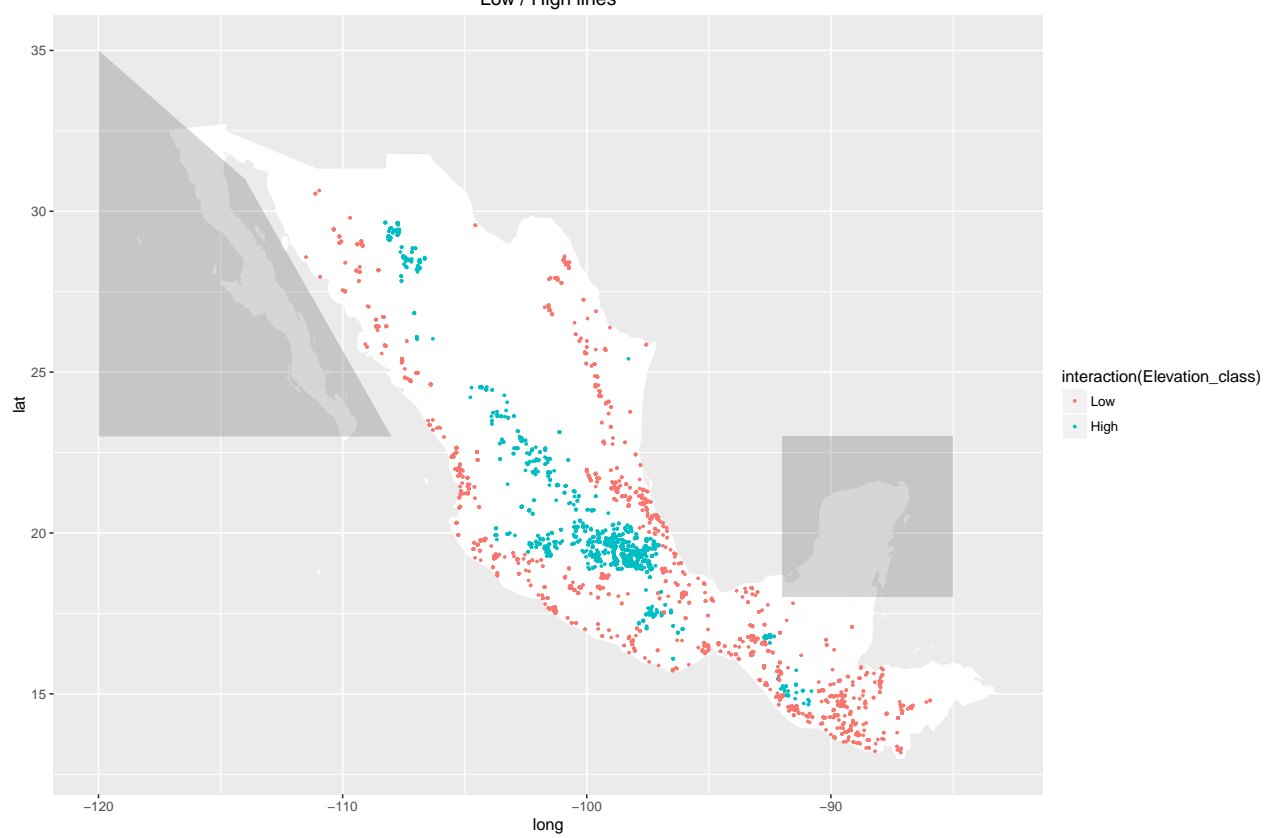
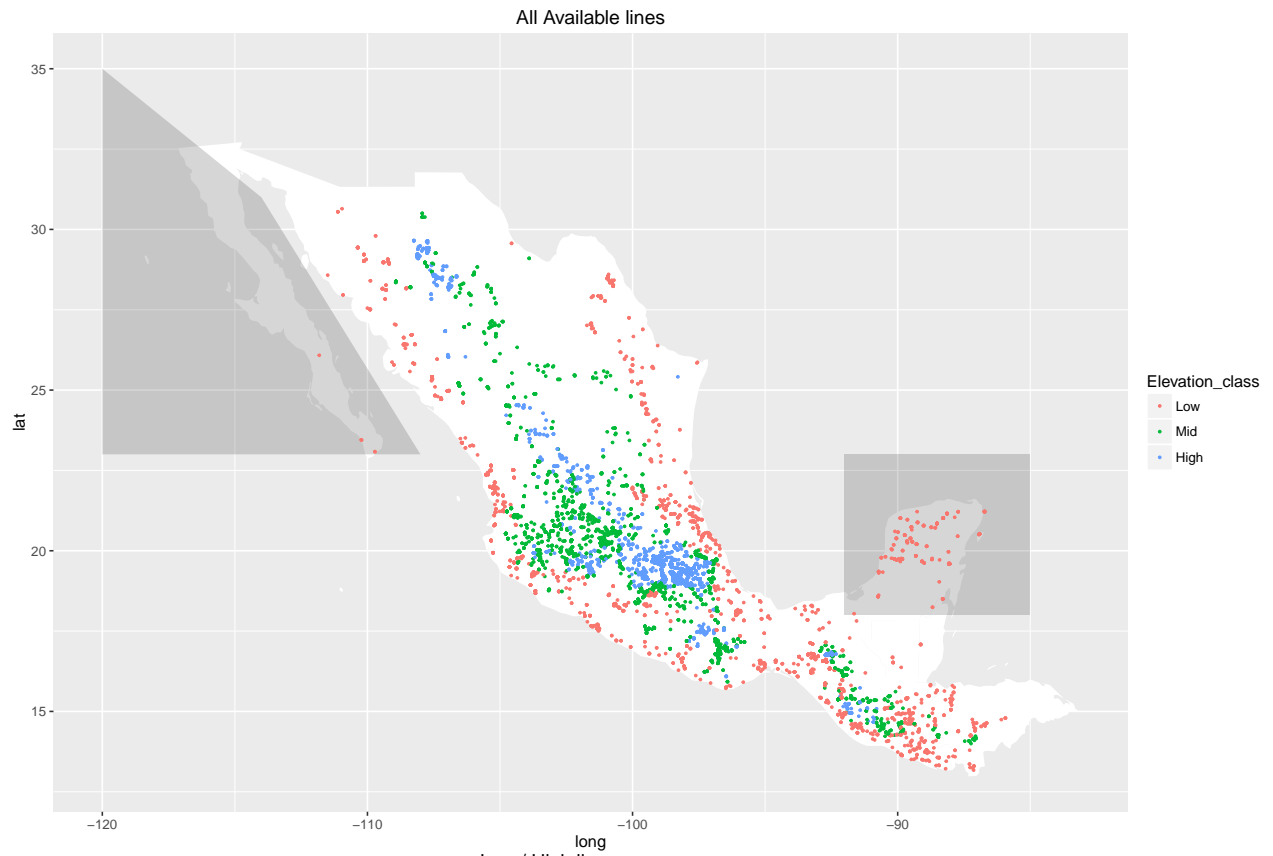
Divide lines by elevation:

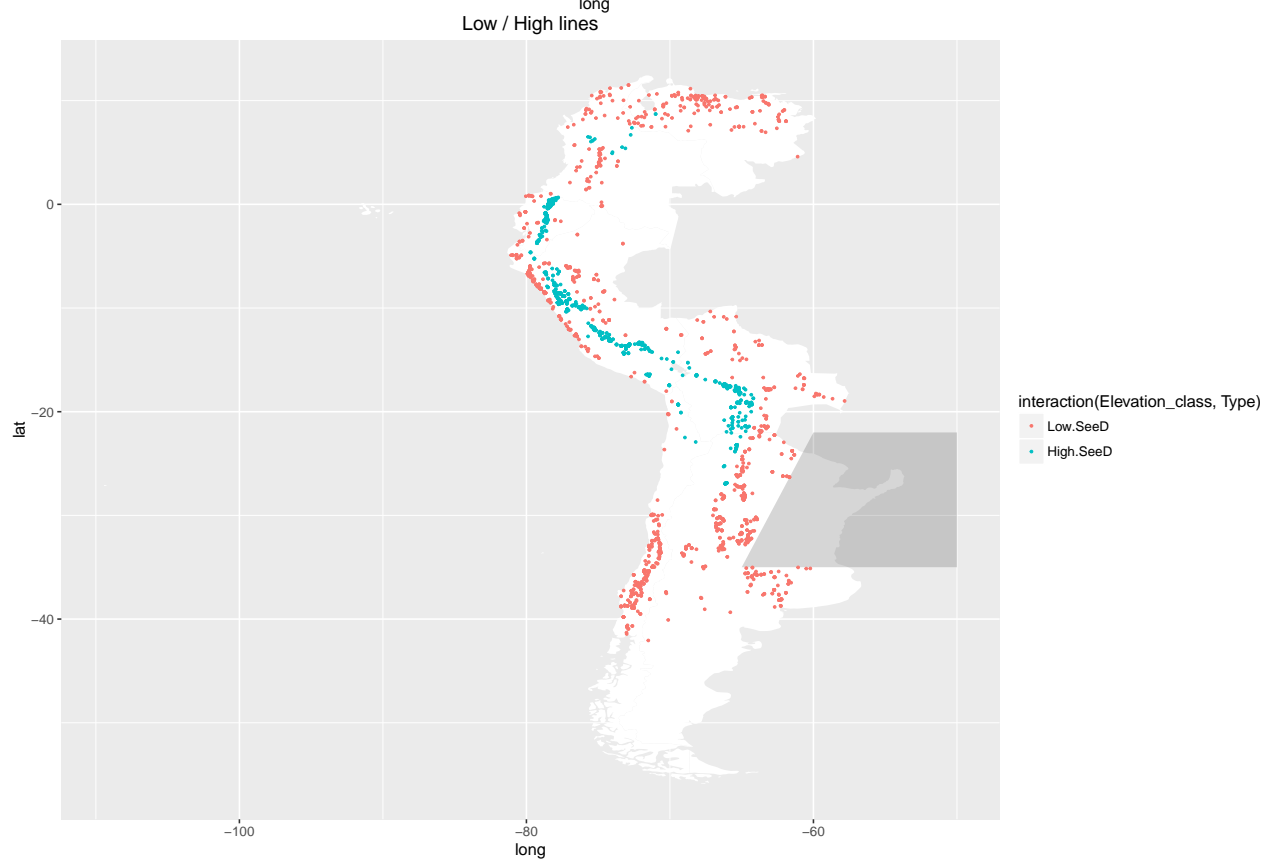
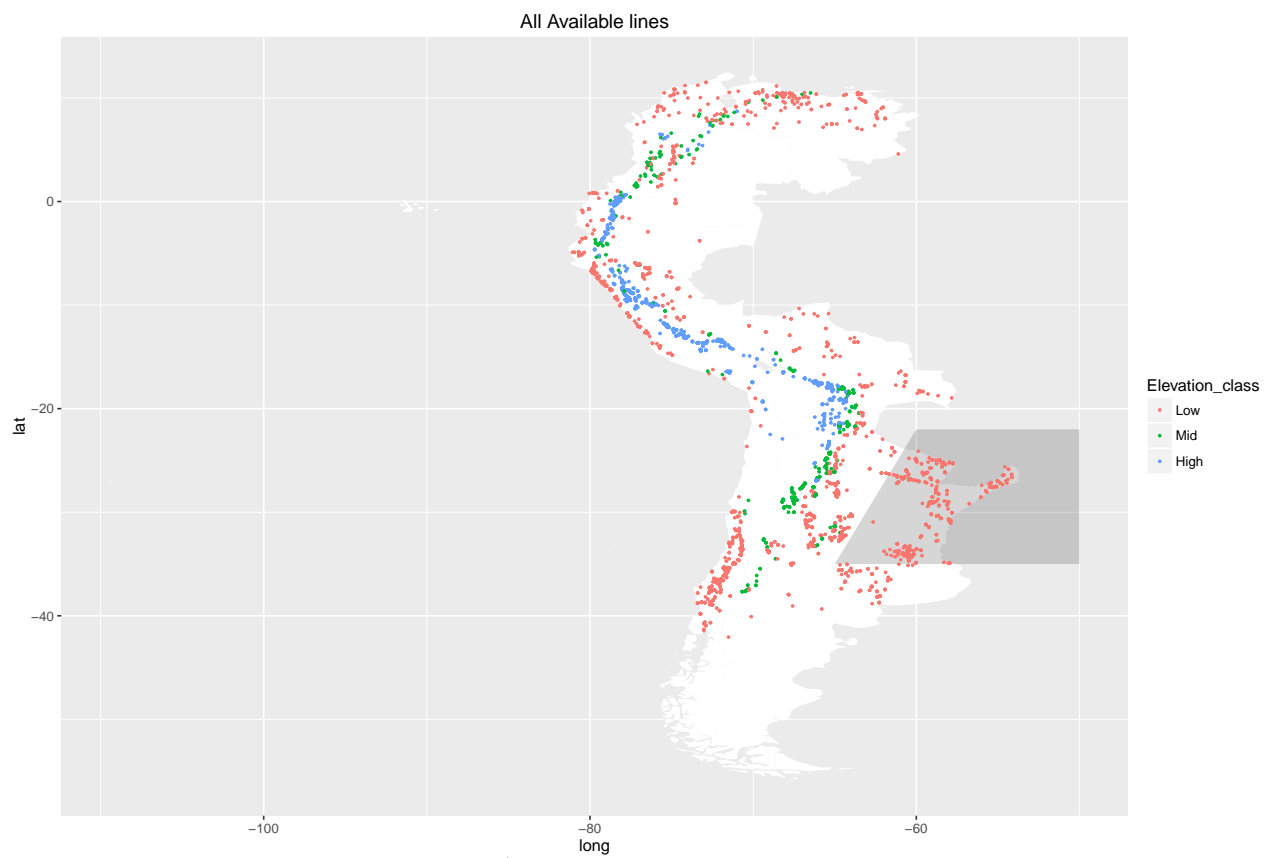
Subset lines by regions (Central America vs South America). Plot

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)





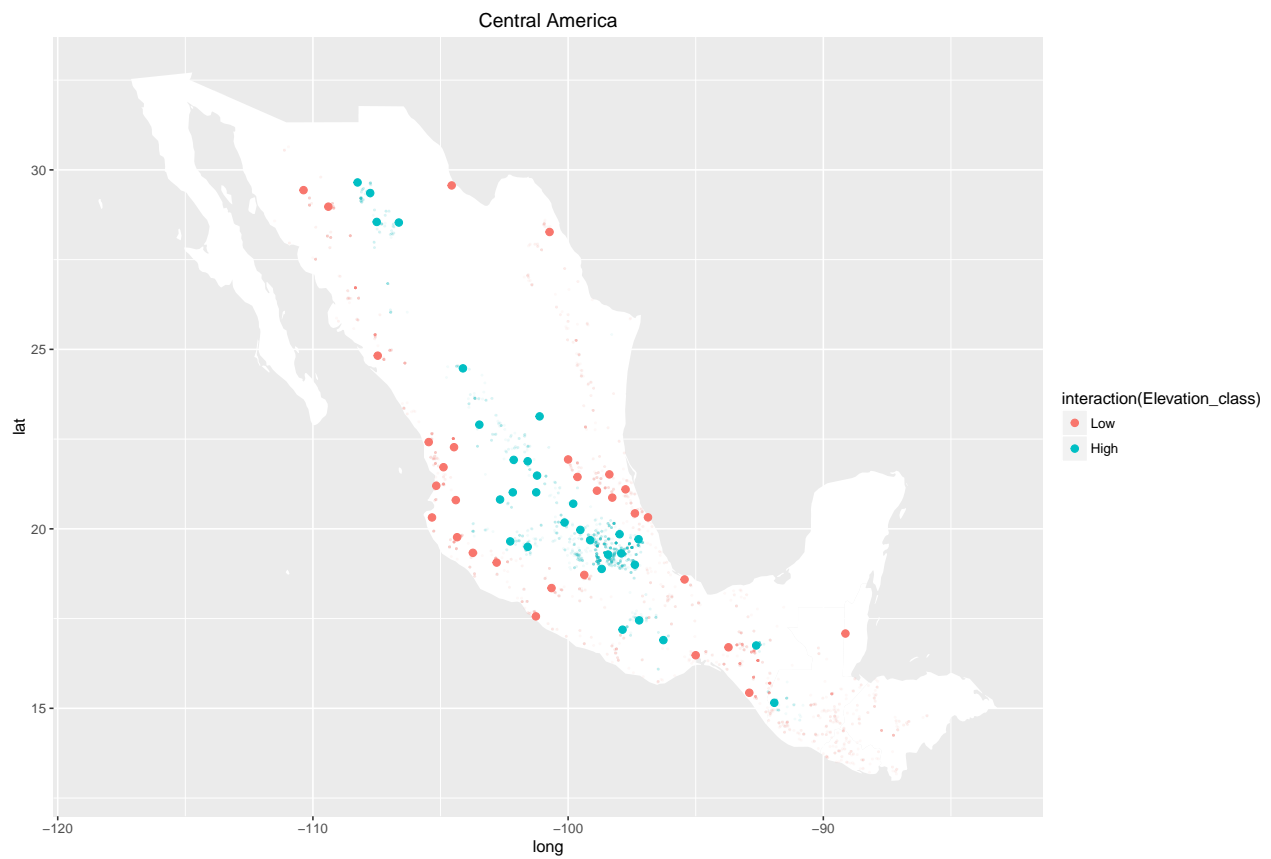
Selection algorithm 1

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
```

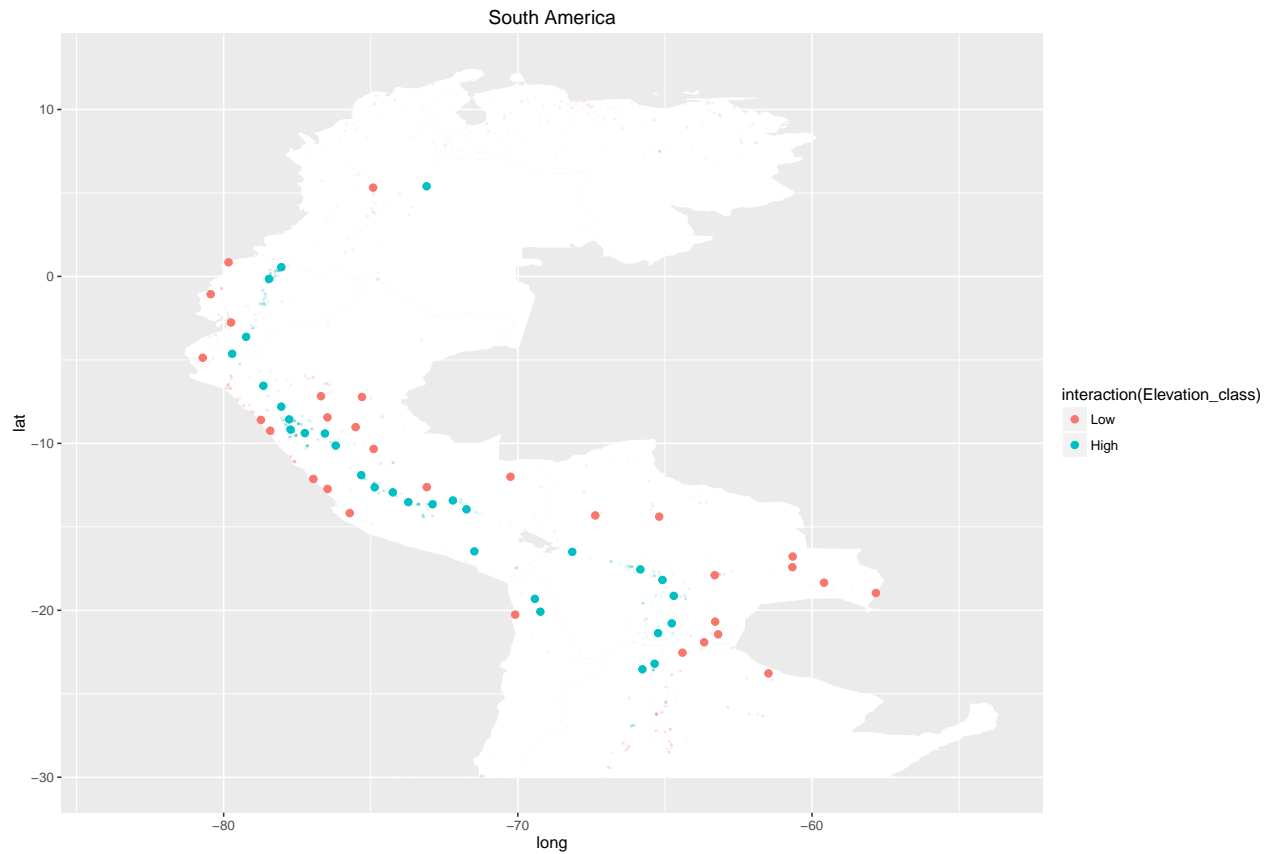
```
## Low  Mid High
##  30   0  30
```



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

```

Now, must try to replace lines that were not successfully ordered

Do selection. For each line that wasn't successfully ordered, select a new one that follows all the old characteristics: with 1 degree of Lat of other line of pair, at least 50 km from every other selected line.

```

## 17410
## 0.15
## 10799
## 0.114607
## 8597
## -0.303652
## [1] CIMMYTMA-007876 CIMMYTMA-028370
## 17144 Levels: CIMMYTMA-000003 CIMMYTMA-000005 ... RIMMA1007
## 16717
## 0.606214
## 7385
## -0.894987
## 8531
## 0.135794
## 14913
## 0.709153
## 8767
## 0.7671107

## pdf
## 2

```

Now save these to a file, with normal info:

```
##      Mex_1 Mex_10 Mex_11 Mex_12 Mex_13 Mex_14 Mex_15 Mex_16 Mex_17 Mex_18
## [1,]      2      2      2      2      2      2      2      2      2      2
##      Mex_19 Mex_2 Mex_20 Mex_21 Mex_22 Mex_23 Mex_24 Mex_25 Mex_26 Mex_27
## [1,]      2      2      2      2      2      2      2      2      2      2
##      Mex_28 Mex_29 Mex_3 Mex_30 Mex_4 Mex_5 Mex_6 Mex_7 Mex_8 Mex_9 SA_1
## [1,]      2      2      2      2      2      2      2      2      2      2
##      SA_10 SA_11 SA_31 SA_13 SA_14 SA_15 SA_16 SA_17 SA_18 SA_19 SA_2
## [1,]      2      2      2      2      2      2      2      2      2      2
##      SA_20 SA_21 SA_22 SA_23 SA_24 SA_25 SA_26 SA_27 SA_28 SA_29 SA_3
## [1,]      2      2      2      2      2      2      2      2      2      2
##      SA_30 SA_4 SA_5 SA_6 SA_7 SA_8 SA_9
## [1,]      2      2      2      2      2      2      2
```

```
## [1] ID
## [2] Long
## [3] Lat
## [4] Elevation
## [5] Race
## [6] GBS
## [7] Elevation_class
## [8] id
## [9] general_identififier
## [10] bank_number
## [11] collnumb
## [12] colldate
## [13] Sample.ID.of.DNA.from.single.plants.used.in.GWAS
## [14] CIMMYT_ID
## [15] SeeD_Elevation
## [16] Elevation_flag
## [17] Pair
## [18] Replaced_line
## <0 rows> (or 0-length row.names)
```

```
##      ID      Long      Lat Elevation  Race  GBS
## 10 CIMMYTMA-002345 -96.70814 19.855393      372 TUXPEN FALSE
## 13 CIMMYTMA-005266 -91.43333 15.283333      2071 OLOTON FALSE
## 25 CIMMYTMA-007876 -79.27404 -1.786495      98      FALSE
## 53 CIMMYTMA-014311 -64.12676 -21.550444      920 APERLA FALSE
## 56 CIMMYTMA-014453 -69.86586 -19.010405      700 HARTAR FALSE
## 67 CIMMYTMA-016818 -75.21928 -14.660126      236      FALSE
## 97 CIMMYTMA-024848 -66.21553 -20.816261      3937 KELLU FALSE
## 101 CIMMYTMA-024949 -70.05131 -17.450516      3206      FALSE
## 115 CIMMYTMA-028370 -78.55979 -1.180281      2753      FALSE
##      Elevation_class      id general_identififier      bank_number
## 10      Low 9686      244939 CIMMYTMA-BANK-002345
## 13      High 11769      247022 CIMMYTMA-BANK-005266
## 25      Low 13705      248958 CIMMYTMA-BANK-007876
## 53      Low 18078      253331 CIMMYTMA-BANK-012151
## 56      Low 18196      253449 CIMMYTMA-BANK-012304
## 67      Low 20288      255541 CIMMYTMA-BANK-015116
## 97      High 26430      261683 CIMMYTMA-BANK-020923
## 101     High 26531      261784 CIMMYTMA-BANK-021544
## 115     High 29316      264569 CIMMYTMA-BANK-026597
##      collnumb      colldate Sample.ID.of.DNA.from.single.plants.used.in.GWAS
```



```

## 10    VERA118 1952-02-07
## 13    GUAT434 1952-01-01
## 25    ECUA323 1954-01-01
## 53    BOZM0296 1977-10-16
## 56    CHZM01008 1981-03-07
## 67    ICA42 1980-02-01
## 97    BOZM1458
## 101    TAC.41 1978-08-01
## 115    ECUA12675
##          CIMMYT_ID SeeD_Elevation Elevation_flag   Pair   Replaced_line
## 10    CIMMYTMA 2345             372             FALSE Mex_26 CIMMYTMA-029313
## 13    CIMMYTMA 5266             2071             FALSE Mex_21 CIMMYTMA-027739
## 25    CIMMYTMA 7876              98             FALSE SA_31 CIMMYTMA-016977
## 53    CIMMYTMA 14311            1177             FALSE SA_30 CIMMYTMA-014314
## 56    CIMMYTMA 14453             699             FALSE SA_10 CIMMYTMA-014470
## 67    CIMMYTMA 16818            236             FALSE SA_27 CIMMYTMA-016810
## 97    CIMMYTMA 24848            3937             FALSE SA_25 CIMMYTMA-014466
## 101    CIMMYTMA 24949            3206             FALSE SA_15 CIMMYTMA-014242
## 115    CIMMYTMA 28370            2753             FALSE SA_31 CIMMYTMA-008156

```