Playing with the Ant Picnic dataset

Jose R. Martínez Batlle

August, 2018

Ant Picnic

- Project leader: Dr. Magdalena Sorger, https://theantlife.com/ and Kristin Bedell
- Project URL: http://studentsdiscover.org/lesson/ant-picnic/
- Data analysis page: http://studentsdiscover.org/lesson/ant-picnic-data-analysis/
- Source data: https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=35710

Packages and scripts

Packages

```
library(dplyr)  #Organize data
library(ez)  #ezCor function, compute an plot correlation matrix
library(devtools)  #For R packages, R scripts
library(gclus)  #For order.single function
library(ggplot2)  #For plots
library(tidyr)  #For data organizing
```

Scripts

```
scripts.others <- "panelutils.R"  #Correlation coefficients upper panel
ghsource.others <- "https://raw.githubusercontent.com/JoeyBernhardt/NumericalEcology/master/"
invisible(source_url(paste0(ghsource.others, scripts.others)))
```

Reading the data

```
#Reading source data. CSV file downloaded from
#https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=35710
tdir <- tempdir()
setwd(tdir)
ghsource <- 'https://raw.githubusercontent.com/geofis/antpicnic/master/'
fname <- 'antpicnic.csv'
download.file(paste0(ghsource, 'sampledata/', fname), fname)
d <- read.csv(fname)
str(d)
```

```
## 'data.frame':    1008 obs. of  18 variables:
##  $ Country                  : Factor w/ 23 levels "Argentina","Australia",..: 10 10 10 10 10 10 10 
##  $ Biome                    : Factor w/ 6 levels "Desert","Grassland/Cold desert",..: 6 6 6 6 6 6 6
##  $ Bait_Type                : Factor w/ 6 levels "Amino Acids",..: 1 1 1 1 1 1 1 1 1 2 ...
##  $ TotalAntBaitType         : int  2 2 2 2 2 2 2 2 2 139 ...
##  $ newAttr                  : logi  NA NA NA NA NA NA ...
##  $ Site_ID                  : int  278 282 279 283 281 286 280 284 285 278 ...
```

```
##  $ Latitude                  : num  63.1 65.4 63.1 59.8 65.4 ...
##  $ Longitude                 : num  25.8 29.6 25.8 23.2 29.6 ...
##  $ Date                      : logi  NA NA NA NA NA NA ...
##  $ Habitat_type              : Factor w/ 119 levels "","3-4 m in the forest (near edge) along a side
##  $ Elevation                 : num  155 260 115 10 222 16 140 50 50 155 ...
##  $ Distance_from_shore       : int  152 202 153 4 204 1 150 25 2 152 ...
##  $ Precipitation             : int  589 588 590 604 585 665 573 672 676 589 ...
##  $ Temperature               : num  2.5 -0.1 2.4 5.3 0 5.2 1.7 5.2 4.8 2.5 ...
##  $ Contributor               : Factor w/ 35 levels "","A.E. Mart?nez Bauer; G.C. Mart?nez",..: 31 31
##  $ Site_Details              : Factor w/ 91 levels "","1/1/2001",..: 2 3 3 40 42 66 78 89 91 2 ...
##  $ Number_of_Ants            : int  0 0 0 0 1 0 0 1 0 4 ...
##  $ Percent_of_Baits_with_Ants: int  0 0 0 0 20 0 0 20 0 40 ...
```

```
N <- nrow(d)
set.seed(131)
d[sample(1:N,10),1:4]
```

```
##       Country                            Biome   Bait_Type
## 209       USA        Temperate seasonal forest Amino Acids
## 126 Australia                           Desert       Sugar
## 296       USA        Temperate seasonal forest        Salt
## 378       USA               Woodland/Shrubland       Water
## 850  Tanzania Tropical seasonal forest/Savanna        Salt
## 531   Ecuador        Temperate seasonal forest   Olive Oil
## 520   Ecuador        Temperate seasonal forest       Sugar
## 239       USA               Woodland/Shrubland Amino Acids
## 327       USA               Woodland/Shrubland      Cookie
## 915  Cambodia Tropical seasonal forest/Savanna       Water
##     TotalAntBaitType
## 209               21
## 126              646
## 296              275
## 378               78
## 850              486
## 531              278
## 520              101
## 239               73
## 327              944
## 915              104
```

Selected variables

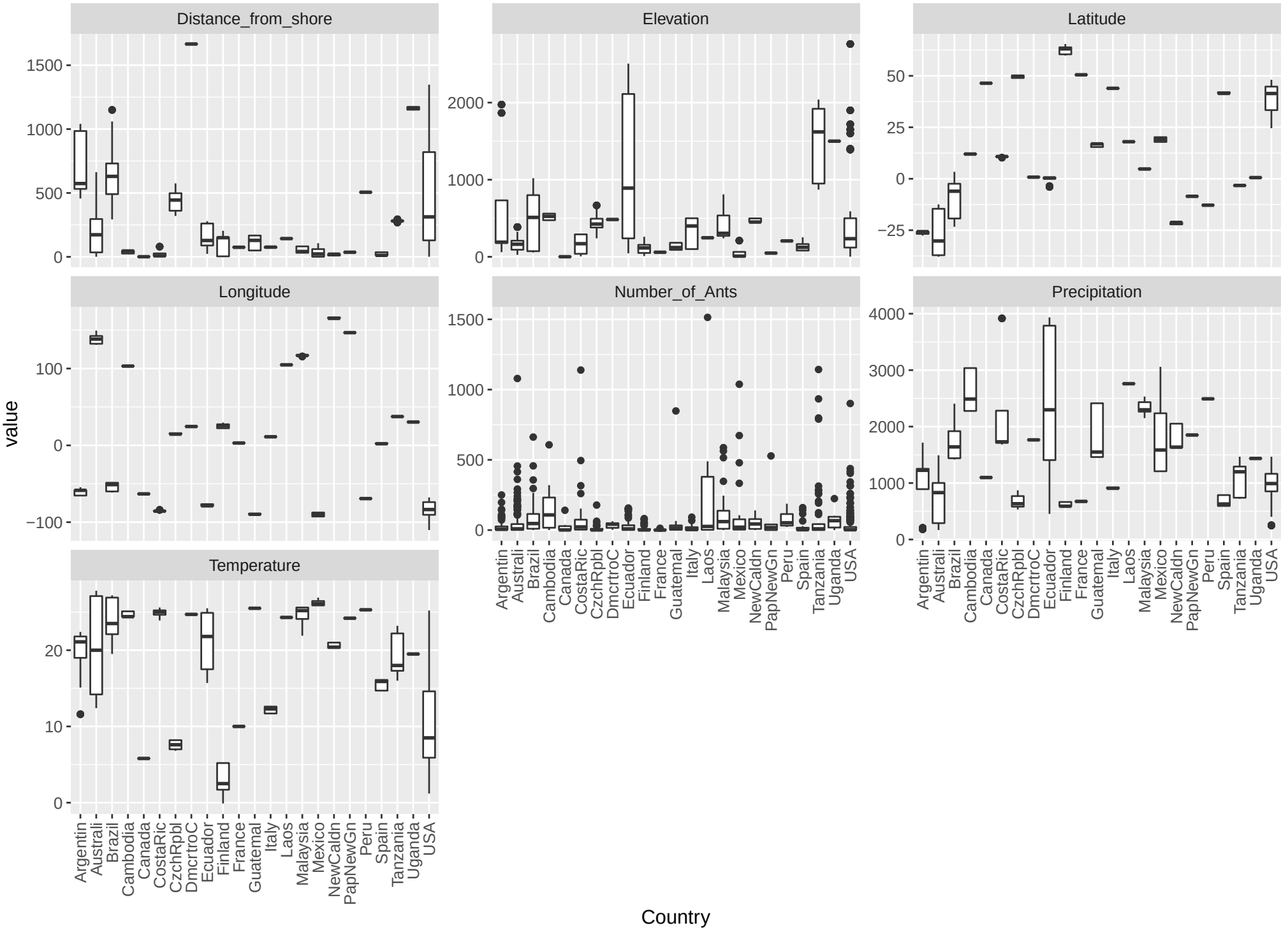
```
d.sel <- d %>% select(
  Country:Bait_Type,
  Latitude:Temperature,
  -Habitat_type,
  -Date,
  Number_of_Ants)
set.seed(131)
d.sel[sample(1:N,10),1:4]
```

```
##       Country                            Biome   Bait_Type   Latitude
## 209       USA        Temperate seasonal forest Amino Acids  33.340000
## 126 Australia                           Desert       Sugar -16.250000
## 296       USA        Temperate seasonal forest        Salt  33.340000
## 378       USA               Woodland/Shrubland       Water  48.090100
```

```
## 850  Tanzania Tropical seasonal forest/Savanna        Salt  -3.330329
## 531   Ecuador        Temperate seasonal forest   Olive Oil   0.326000
## 520   Ecuador        Temperate seasonal forest       Sugar   0.319000
## 239       USA               Woodland/Shrubland Amino Acids  40.128694
## 327       USA               Woodland/Shrubland      Cookie  45.959389
## 915  Cambodia Tropical seasonal forest/Savanna       Water  12.004710
```

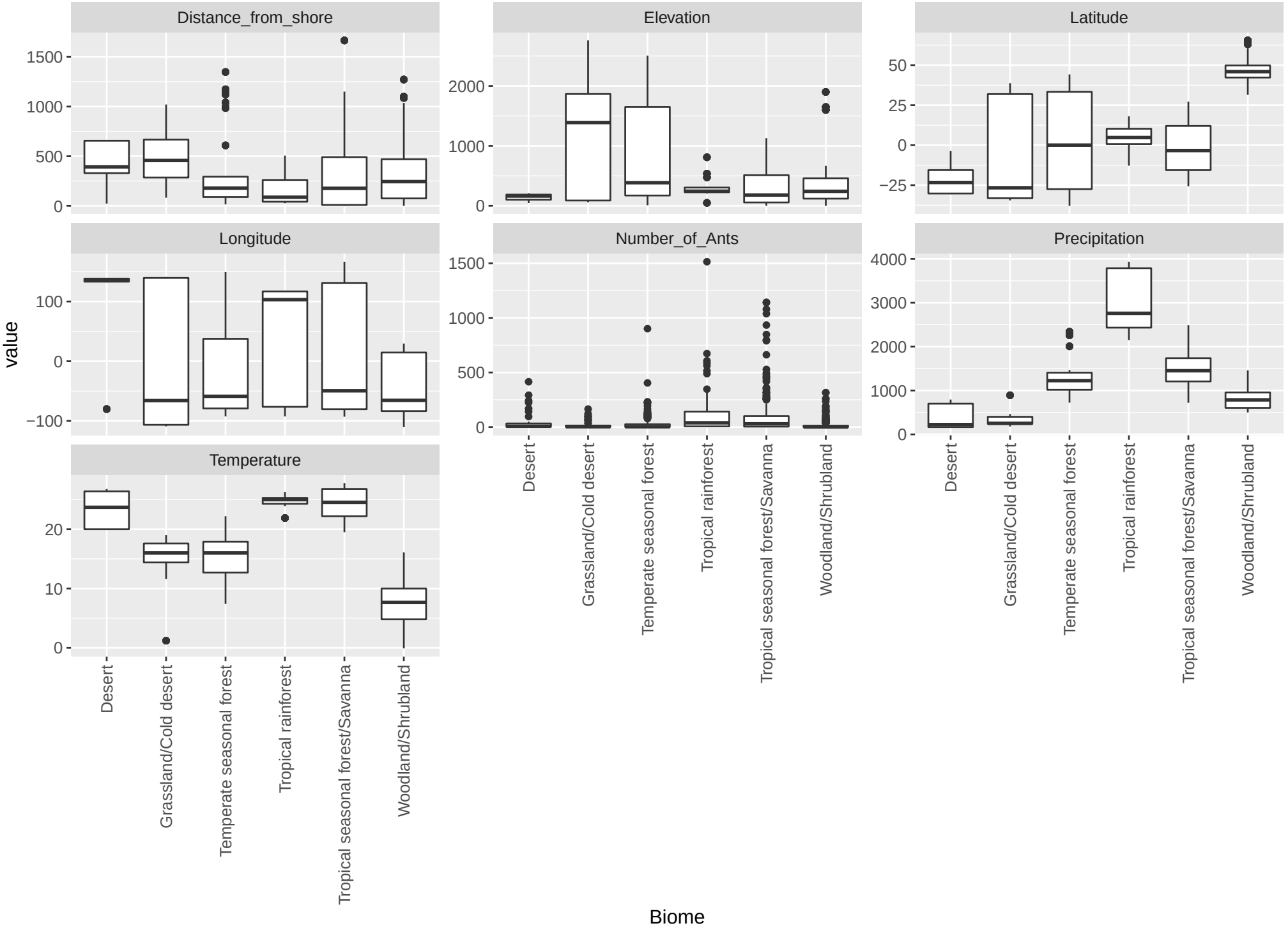
Environmental variables and number of ants, by country

```
basics.country.bp <- d.sel %>% gather(variable,value,-Country,-Biome,-Bait_Type) %>%
  mutate(Country=abbreviate(Country,8)) %>%
  ggplot(aes(x=Country, y=value)) +
  geom_boxplot() +
  theme(
    text = element_text(size = 12),
    axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)
    ) +
  facet_wrap(variable~., scales = 'free_y', nrow=3)
```


Distance_from_shore
Elevation
Latitude
Longitude
Number_of_Ants
Precipitation
Temperature
value
Country
Argentin
Australi
Brazil
Cambodia
Canada
CostaRic
CzchRpbl
DmcrtroC
Ecuador
Finland
France
Guatemal
Italy
Laos
Malaysia
Mexico
NewCaldn
PapNewGn
Peru
Spain
Tanzania
Uganda
USA

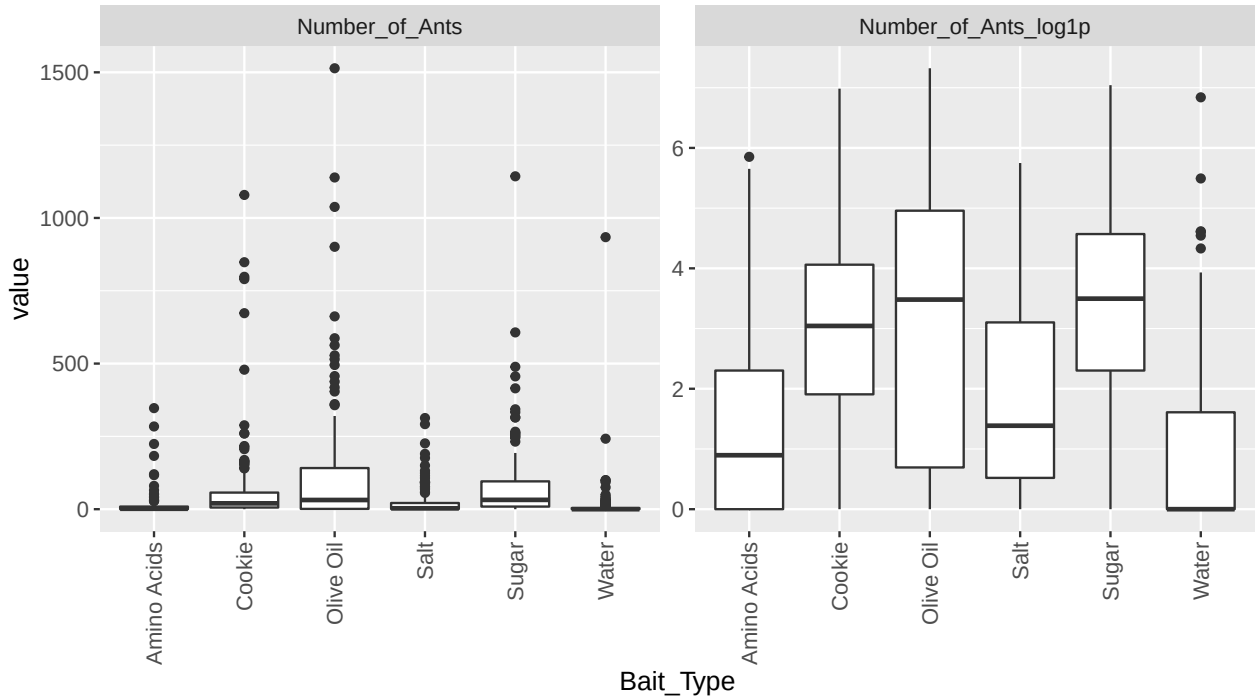
Environmental variables and number of ants, by biome

```
basics.biome.bp <- d.sel %>% gather(variable,value,-Country,-Biome,-Bait_Type) %>%
  ggplot(aes(x=Biome, y=value)) +
  geom_boxplot() +
  theme(
    text = element_text(size = 12),
    axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)
    ) +
  facet_wrap(variable~., scales = 'free_y', nrow=3)
```


Distance_from_shore
Elevation
Latitude
Longitude
Number_of_Ants
Precipitation
Temperature
value
Biome
Desert
Grassland/Cold desert
Temperate seasonal forest
Tropical rainforest
Tropical seasonal forest/Savanna
Woodland/Shrubland

Number of ants, by bait type

```
basics.bait.bp <- d.sel %>%
  select(Bait_Type, Number_of_Ants) %>%
  mutate(Number_of_Ants_log1p = log1p(Number_of_Ants)) %>%
  gather(variable,value,-Bait_Type) %>%
  ggplot(aes(x=Bait_Type, y=value)) +
  geom_boxplot() +
  theme(
    text = element_text(size = 12),
    axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)
    ) +
  facet_wrap(variable~., scales = 'free_y', ncol = 2)
basics.bait.bp
```



Data frame for correlation analysis

```
d.fcor <- d.sel %>% select_if(is.numeric)
set.seed(131)
d.fcor[sample(1:N,10),1:5]
```

```
##       Latitude Longitude Elevation Distance_from_shore Precipitation
## 209  33.340000 -81.54000       100                 177          1205
## 126 -16.250000 133.36670       211                 420           700
## 296  33.340000 -81.54000       100                 177          1205
## 378  48.090100 -90.82438       500                 816           721
## 850  -3.330329  37.64310      1130                 281           948
## 531   0.326000 -78.94900       839                 119          2338
## 520   0.319000 -78.95300       943                 120          2338
## 239  40.128694 -88.12194       210                1037           951
## 327  45.959389 -68.36153       157                 135           999
## 915  12.004710 103.20240       560                  46          2488
```

```
d.fcor.o <- order.single(cor(d.fcor)) #Ordered according to r value
```

Data frame for correlation analysis with transformed variables

```
d.fcor.l <- d.fcor %>%
  mutate(
    Latitude=abs(Latitude),
    Longitude=abs(Longitude)) %>%
  mutate_all(., funs(log1p))
set.seed(131)
d.fcor.l[sample(1:N,10),1:5]
```

```
##      Latitude Longitude Elevation Distance_from_shore Precipitation
## 209 3.5363109  4.413283  4.615121            5.181784      7.095064
## 126 2.8478121  4.900573  5.356586            6.042633      6.552508
## 296 3.5363109  4.413283  4.615121            5.181784      7.095064
## 378 3.8936574  4.519878  6.216606            6.705639      6.582025
## 850 1.4656435  3.654368  7.030857            5.641907      6.855409
## 531 0.2821669  4.381389  6.733402            4.787492      7.757479
## 520 0.2768739  4.381439  6.850126            4.795791      7.757479
## 239 3.7167060  4.490006  5.351858            6.945051      6.858565
## 327 3.8492832  4.239332  5.062595            4.912655      6.907755
## 915 2.5653116  4.646335  6.329721            3.850148      7.819636
```

```
d.fcor.l.o <- order.single(cor(d.fcor.l))
```

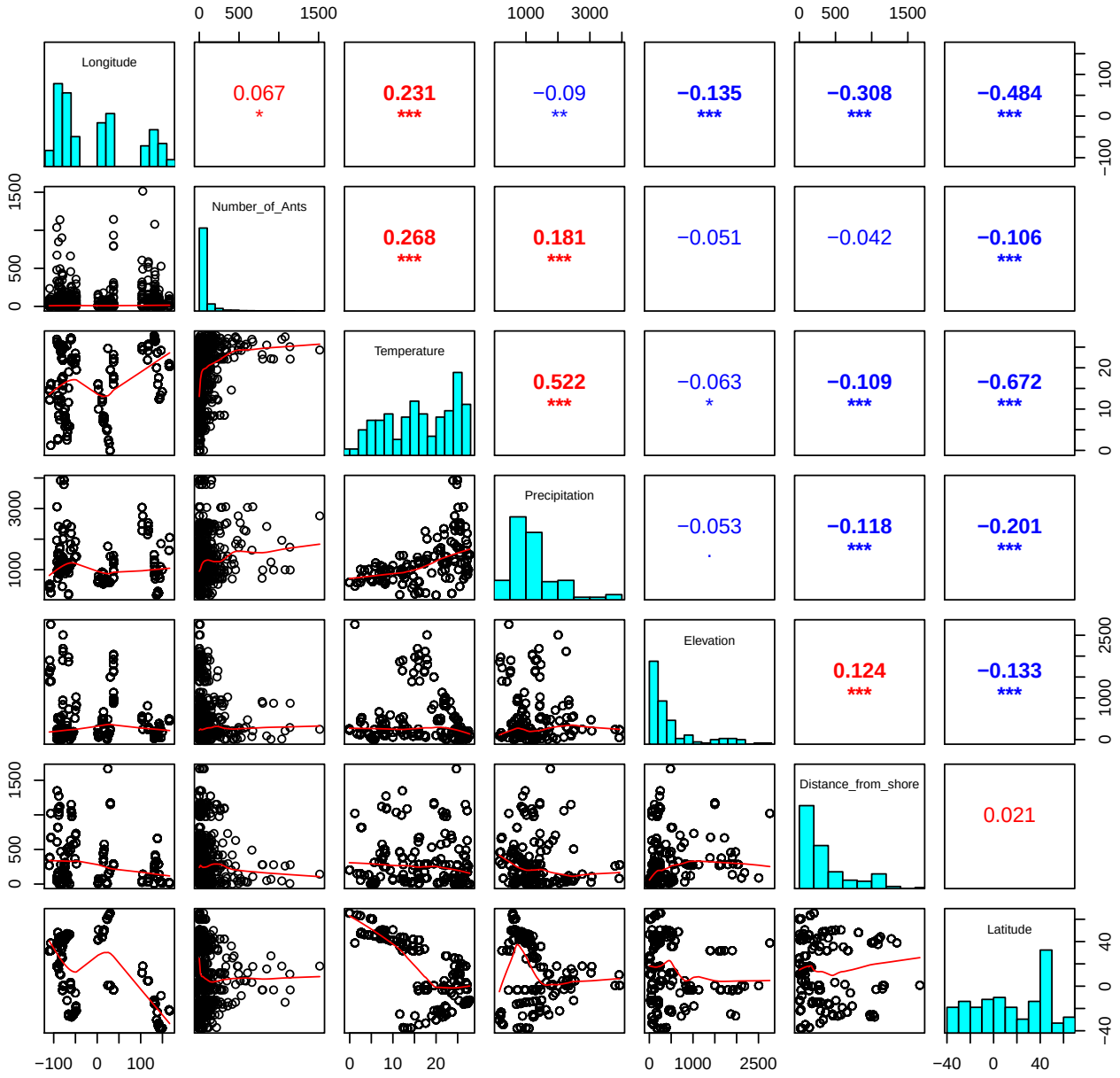
Correlation panel using ez package

```
ezCor(
  d.fcor,
  r_size_lims = c(4,9),
  label_size = 3) +
  labs(title="Environmental variables correlation panel")
```

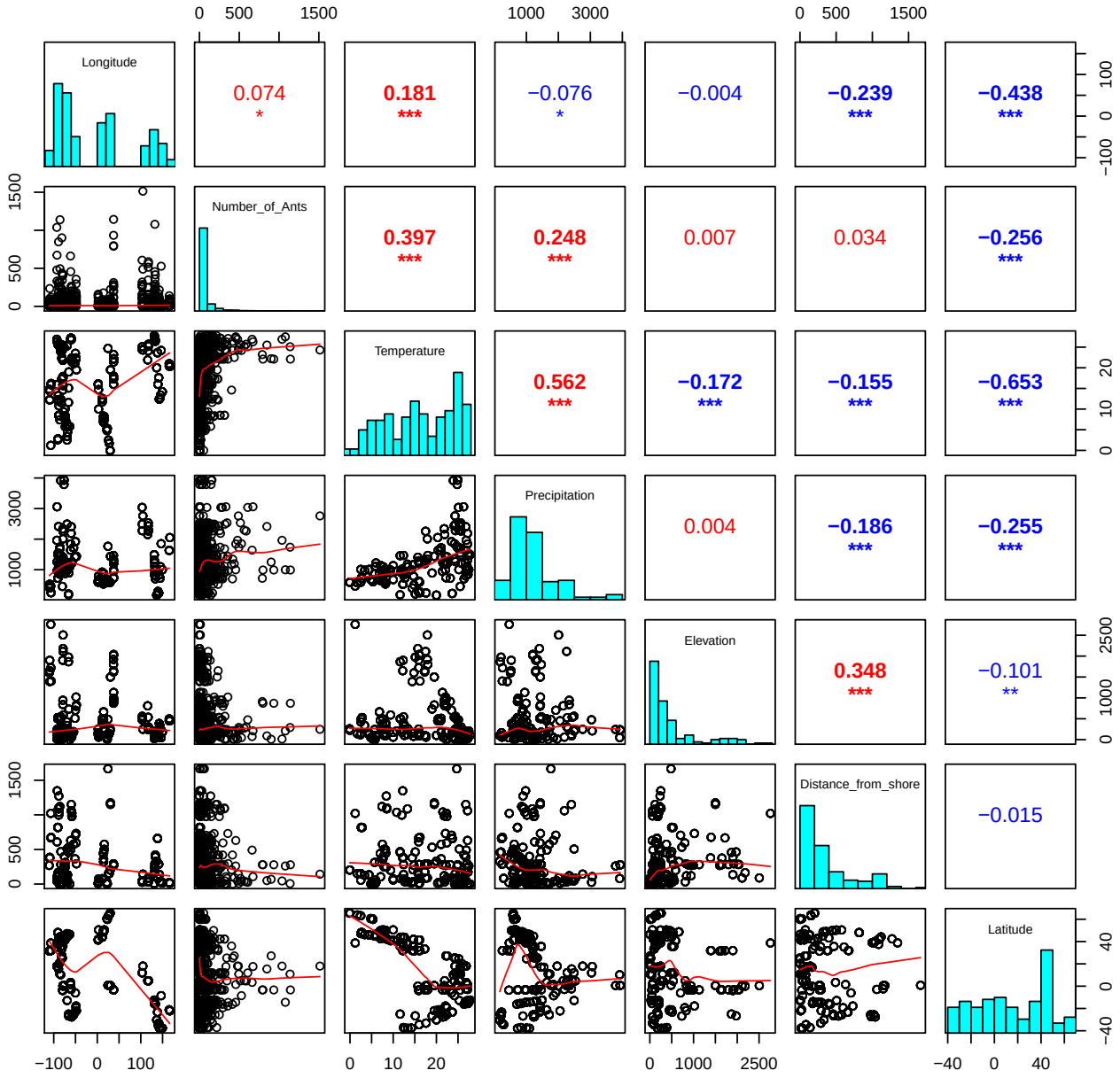
Environmental variables correlation panel

Correlation panel of raw variables using pairs function and panelutils.R script

```
#Pearson
pairs(d.fcor[,d.fcor.o],
      lower.panel = panel.smooth,
      upper.panel = panel.cor,
      diag.panel = panel.hist,
      main = "Pearson Correlation Panel")
```

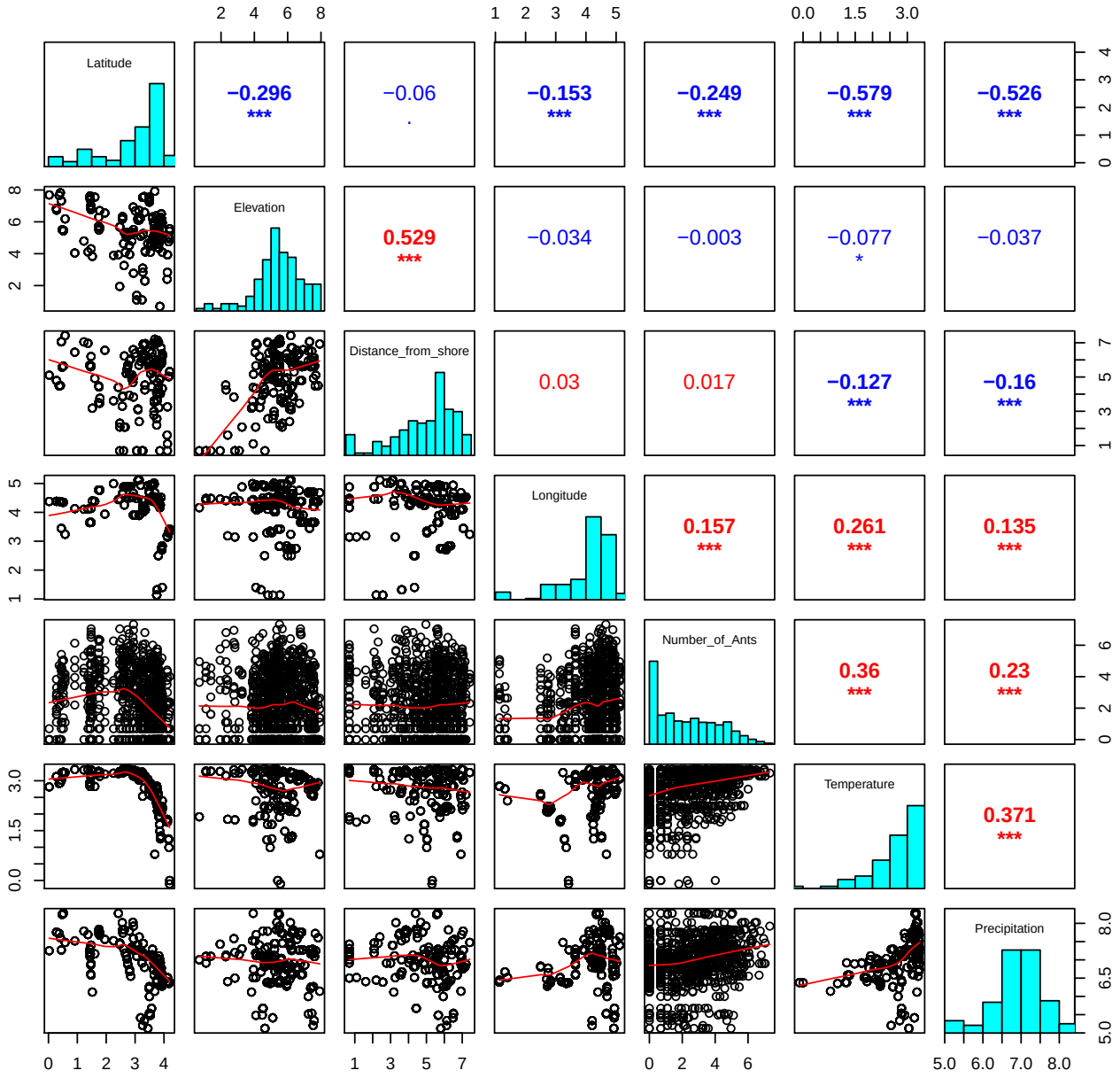



```
#Spearman
pairs(d.fcor[,d.fcor.o],
      lower.panel = panel.smooth,
      upper.panel = panel.cor,
      diag.panel = panel.hist,
      method = 'spearman',
      main = "Spearman Correlation Panel")
```

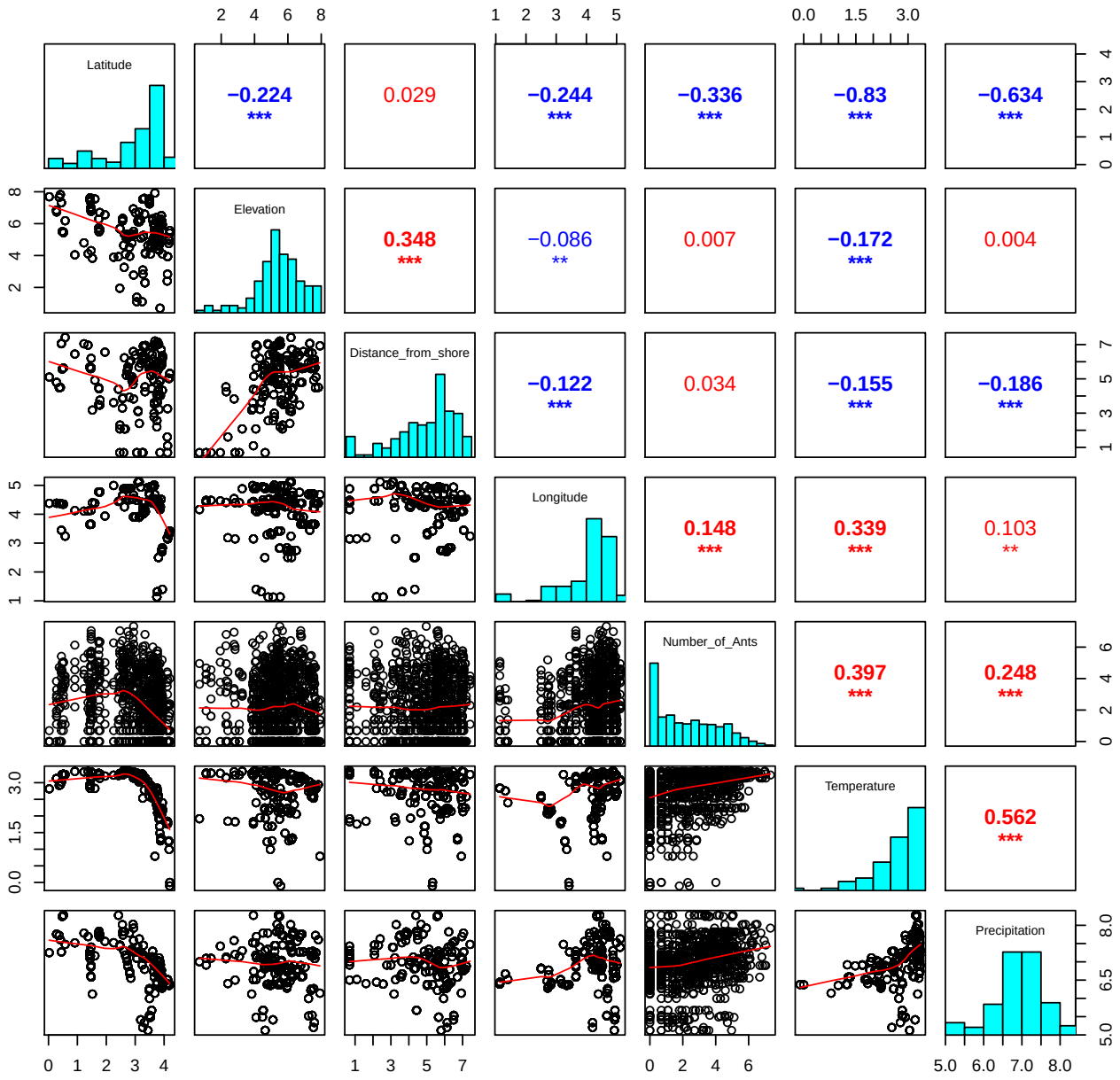


Correlation panel of log-transformed variables using pairs function and panelutils.R script

```
#Pearson
pairs(d.fcor.l[,d.fcor.l.o],
  lower.panel = panel.smooth,
  upper.panel = panel.cor,
  diag.panel = panel.hist,
  main = "Pearson Correlation Panel (log1p)")
```




```
#Spearman
pairs(d.fcor.l[,d.fcor.l.o],
      lower.panel = panel.smooth,
      upper.panel = panel.cor,
      diag.panel = panel.hist,
      method = 'spearman',
      main = "Spearman Correlation Panel (log1p)")
```


Spearman Correlation Panel (log1p)
Latitude
Elevation
Distance_from_shore
Longitude
Number_of_Ants
Temperature
Precipitation
−0.224 ***
0.029
−0.244 ***
−0.336 ***
−0.83 ***
−0.634 ***
0.348 ***
−0.086 **
0.007
−0.172 ***
0.004
−0.122 ***
0.034
−0.155 ***
−0.186 ***
0.148 ***
0.339 ***
0.103 **
0.397 ***
0.248 ***
0.562 ***