

## R-scripts used in data analysis

```
# representation of demographic data of respondents

require(grDevices)
pie(rep(1, 24), col = rainbow(24), radius = 0.9)

par(mfrow=c(2,2))

pie.gender <- c(37,63)
names(pie.gender) <- c("female-37%", "male-63%")
pie(pie.gender,col=c("red", "blue"),main="distribution of taxonomists per
gender")

pie.age <- c(21,68,11)
names(pie.age) <- c("young-21%", "middle age-68%", "mature-11%")
pie(pie.age,col=c("red", "blue", "green"),main="distribution of taxonomists per
age")

pie.acad.status <- c(32,26,21,5,16)
names(pie.acad.status) <- c("Prof-32%", "Dr-26%", "Ass-Prof-21%", "Diploma-
5%", "MSc-16%")
pie(pie.acad.status,col=c("red", "blue", "green", "purple", "yellow"),main =
"academic status/ranks of taxonomists")

pie.primary_affiliation <- c(32,68)
names(pie.primary_affiliation) <- c("Res. Inst-32%", "Universities-68%")
pie(pie.primary_affiliation,col=c("red", "blue"),main="primary affiliation of
taxonomists")

#####
par(mfrow=c(1,1))

pie.biome <- c(9,7,20.5,20.5,11,16,16,0)
names(pie.biome) <- c("Albany Thicket-9%", "Forest-7%", "Fynbos-
20.5%", "Grassland-20.5%", "Nama Karoo-11%", "Savannah-16%", "Succulent Karoo-
16%", "Indian Ocean Costal Belt-0%")
pie(pie.biome,col=c("red", "blue", "purple", "violetred1", "green3", "cornsilk",
"cyan", "white", "yellow"),main="biome of research interest")

# distribution of taxonomists per year of experience
pie.year.exp <- c(16,42,37,5)
names(pie.year.exp) <- c("<10 (16%)", "10-20 (42%)", "20-30 (37%)", ">30 (5%)")
pie(pie.year.exp,col=c("red", "blue", "purple", "violetred1"),main="years of
experience in taxonomy")
```

```
# distribution of taxonomists per year of experience
pie.year.exp <- c(16,42,37,5)
names(pie.year.exp) <- c("<10 (16%)", "10-20 (42%)", "20-30 (37%)", ">30 (5%)")
pie(pie.year.exp,col=c("red", "blue", "purple", "violetred1"),main="years of
experience in taxonomy")
```

```

# distribution of taxonomists per number of species described
pie.taxon.nber.sp.described <- c(55,22,6,11,6)
names(pie.taxon.nber.sp.described) <- c("1-10sp(55%)", "10-20sp(22%)", "20-30sp(6%)", "30-90sp(11%)", ">90sp(6%)")
pie(pie.taxon.nber.sp.described,col=c("red","blue","purple","violetred1","green3"),main="taxonomists per species described")

# distribution of taxonomists per prediction of unknown species
pie.taxon.predict <- c(25,17,42,8,8)
names(pie.taxon.predict) <- c("100-200sp(25%)", "200-1000sp(17%)", "1000-2000sp(42%)", "2000-3000sp(8%)", ">3000sp(8%)")
pie(pie.taxon.predict,col=c("red","blue","purple","violetred1","green3"),main="experts opinion on unknown species")

# distribution of taxonomists per publication in taxonomy
pie.taxon.paper <- c(32,37,16,5,10)
names(pie.taxon.paper) <- c("1-20papers(32%)", "20-50papers(37%)", "50-80papers(16%)", "80-200papers(5%)", ">200papers(10%)")
pie(pie.taxon.paper,col=c("red","blue","purple","violetred1","green3"),main="taxonomists per publications")

# Prediction of number of unknown species in SA
# Based on expert opinion
data_questionnaire <-
read.table("C:\\Users\\kowiyouy\\Desktop\\UJ\\MSc\\2019\\MSc
supervision\\botany\\DATA\\questionnaire_responses.txt",header = TRUE)
attach(data_questionnaire)
names(data_questionnaire)
mean(data_questionnaire$est_unknown_species)

#####Temporal trend of species description

library(ggplot2)
library(tidyr)
library(dplyr)

data_patience <- read.table("C:\\Users\\kowiyouy\\Desktop\\UJ\\MSc\\2019\\MSc
supervision\\botany\\DATA\\data_trend_real.txt",header=TRUE)

attach(data_patience)
names(data_patience)

#Create basic line plots
# Basic line plot of changes in number of species published over time

ggplot(data = data_patience, aes(x = year, y = number_species),ylab="number
species described")+
  geom_line(color = "#00AFBB", size = 2)

# Basic line plot of changes in number of authors who published species over
time
ggplot(data = data_patience, aes(x = year, y = Number_authors),ylab="number
authors")+

```

```

geom_line(color = "#00AFBB", size = 2)

#Plot multiple time series data
#Here, we'll plot the variables psavert and uempmed by dates.
#You should first reshape the data using the tidyr package: - Collapse psavert
and uempmed values in the same column (new column).
#R function: gather()[tidyr] - Create a grouping variable that with levels =
psavert and uempmed

###create the line plot of both number of species and number of authors in one
graph.

df <- data_patience %>%
  select(year, number_species, Number_authors) %>%
  gather(key = "variable", value = "value", -year)

## Area plot
ggplot(df, aes(x = year, y = value)) +
  geom_area(aes(color = variable, fill = variable),
            alpha = 0.5, position = position_dodge(0.8)) +
  scale_color_manual(values = c("#00AFBB", "#E7B800", "#FC4E07", "#FF9999")) +
  scale_fill_manual(values = c("#00AFBB", "#E7B800", "#FC4E07", "#FF9999"))

```

```

### GARCH model for prediction

library(quantmod)
library(rugarch)

###
model_species <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(0,0)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current
condition

#fit the model
number_species_Garch <- ugarchfit(spec=model_species,data=number_species) #AIC=
10.048
number_species_Predict <- ugarchboot(number_species_Garch,n.ahead = 10,method =
c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_species_Predict,which=2)
###

number_species <- data_patience$number_species

###
model_species <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(2,2)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current
condition

```

```

#fit the model
number_species_Garch <- ugarchfit(spec=model_species,data=number_species) #AIC=
9.5993
number_species_Predict <- ugarchboot(number_species_Garch,n.ahead = 10,method =
c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_species_Predict,which=2)

#####
###
model_species <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(3,3)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current
condition

#fit the model
number_species_Garch <- ugarchfit(spec=model_species,data=number_species) #AIC=
9.5877
number_species_Predict <- ugarchboot(number_species_Garch,n.ahead = 10,method =
c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_species_Predict,which=2)

###
model_species <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(4,4)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current
condition

#fit the model
number_species_Garch <- ugarchfit(spec=model_species,data=number_species) #AIC=
9.6035
number_species_Predict <- ugarchboot(number_species_Garch,n.ahead = 10,method =
c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_species_Predict,which=2)

#####
model_species <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(5,5)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current
condition

#fit the model
number_species_Garch <- ugarchfit(spec=model_species,data=number_species) #AIC=
9.6122
number_species_Predict <- ugarchboot(number_species_Garch,n.ahead = 10,method =
c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_species_Predict,which=2)

#
##best model identified
model_species <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(1,1)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current

```

```
condition
```

```
#fit the model
number_species_Garch <- ugarchfit(spec=model_species,data=number_species) #AIC=
9.5847
number_species_Predict <- ugarchboot(number_species_Garch,n.ahead = 150,method
= c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_species_Predict,which=2,ylab="trend in number species predicted")
abline(h=0,lwd=2,lty="dashed")
abline(v=45,lwd=2,lty="dashed")
#####
```

```
number_species_Predict <- ugarchboot(number_species_Garch,n.ahead = 500,method
= c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_species_Predict,which=2,ylab="trend in number species predicted")
abline(h=0,lwd=2,lty="dashed")
abline(v=30,lwd=2,lty="dashed")
```

```
### predict number of authors
```

```
number_species <- data_patience$number_species
Number_authors <- data_patience$Number_authors
###
model_species <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(2,2)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current
condition
model_authors <- ugarchspec(variance.model =
list(model="sGARCH",garchOrder=c(1,1)),mean.model =
list(armaOrder=c(2,2)),distribution.model="std") #armaOrder=c(1,1) means we
want to use the x for one day before and Y for 1 day before to predict current
condition

#fit the model
number_authors_Garch <- ugarchfit(spec=model_authors,data=Number_authors) #AIC=
6.0371

number_authors_Predict <- ugarchboot(number_authors_Garch,n.ahead = 150,method
= c("Partial","Full")[1])#n.ahead = 10 means we want to predict for 10 days
plot(number_authors_Predict,which=2)
```