

Varietal threat

Natalia Estrada Carmona

08 October 2021

```
rm(list=ls())

library("magrittr") # for piping (%>%)
library(dplyr) # for data manipulation
library(stringr) # string manipulation
library(tidyr)
library(reshape2)
library(Matrix)
library(data.table)
library(metafor) #meta-analysis##
library(Formula)
library(ggplot2)
library(histogram)
library(grid)
library(funModeling)
library(frequency)
library(forestplot)
library(rlang)
library(doParallel)
library(parallel)
library(iterators)
library(foreach)
library(metafor)
library(meta)
library(readxl)
library(openxlsx)
library(dplyr)
library(tidyverse)
library(varhandle)
library(funModeling)
library(EnvStats)
library(ggstatsplot)
library(RColorBrewer)
library(gridExtra)
library(vegan)
library(networkD3)
library(plotly)

devtools::install_github("MathiasHarrer/dmetar")

setwd("C:\\Users\\necarmona\\OneDrive - CGIAR\\trade_offs_india\\Eshan")
```

#Supplemental material data & scripts ### Data sources

```
TableA_gral<-read_excel("C:\\Users\\necarmona\\OneDrive - CGIAR\\trade_offs_india\\Eshan\\original_FGD_07_10_2021_for_dataverse.xlsx",
                        sheet = "TableA_gral", col_names=TRUE)

TableB_status_trends<-read_excel("C:\\Users\\necarmona\\OneDrive - CGIAR\\trade_offs_india\\Eshan\\original_FGD_07_10_2021_for_dataverse.xlsx",
                                  sheet = "TableB_status_trends", col_names=TRUE)

Participants<-read_excel("C:\\Users\\necarmona\\OneDrive - CGIAR\\trade_offs_india\\Eshan\\original_FGD_04_07_2021.xlsx",
                         sheet = "Participants", col_names=TRUE) # data not shared for privacy and data protection reasons

Metadata<-read_excel("C:\\Users\\necarmona\\OneDrive - CGIAR\\trade_offs_india\\Eshan\\original_FGD_07_10_2021_for_dataverse.xlsx",
                    sheet = "Metadata", col_names=TRUE)

crop_spp_var_names<-read_excel("C:\\Users\\necarmona\\OneDrive - CGIAR\\trade_offs_india\\Eshan\\original_FGD_07_10_2021_for_dataverse.xlsx",
                               sheet = "names_checked", col_names=TRUE)
```

###Figure S2 focus group discussion – FGD population characterization across age and sex groups; states and sex groups with total percentage and the n per state/sex group The map was created with ArcGIS

```

Participants_count<-Participants%>%
  group_by(Age, Gender_rev)%>%
  summarise(n=n())%>%
  mutate(per=n*100/sum(n))%>%
  mutate(Sex=if_else(Gender_rev==1, "M", "F"))

pop<-ggplot(Participants_count, aes(x = Age, fill = Sex,
                                   y = ifelse(test = Sex == "M",
                                              yes = -n, no = n))) +
  geom_bar(stat = "identity") +
  labs(y = "Population (no.)") +
  scale_fill_manual(values = c("#B6C7CE", "#7998A5"))+
  theme(legend.position="none")+
  coord_flip()

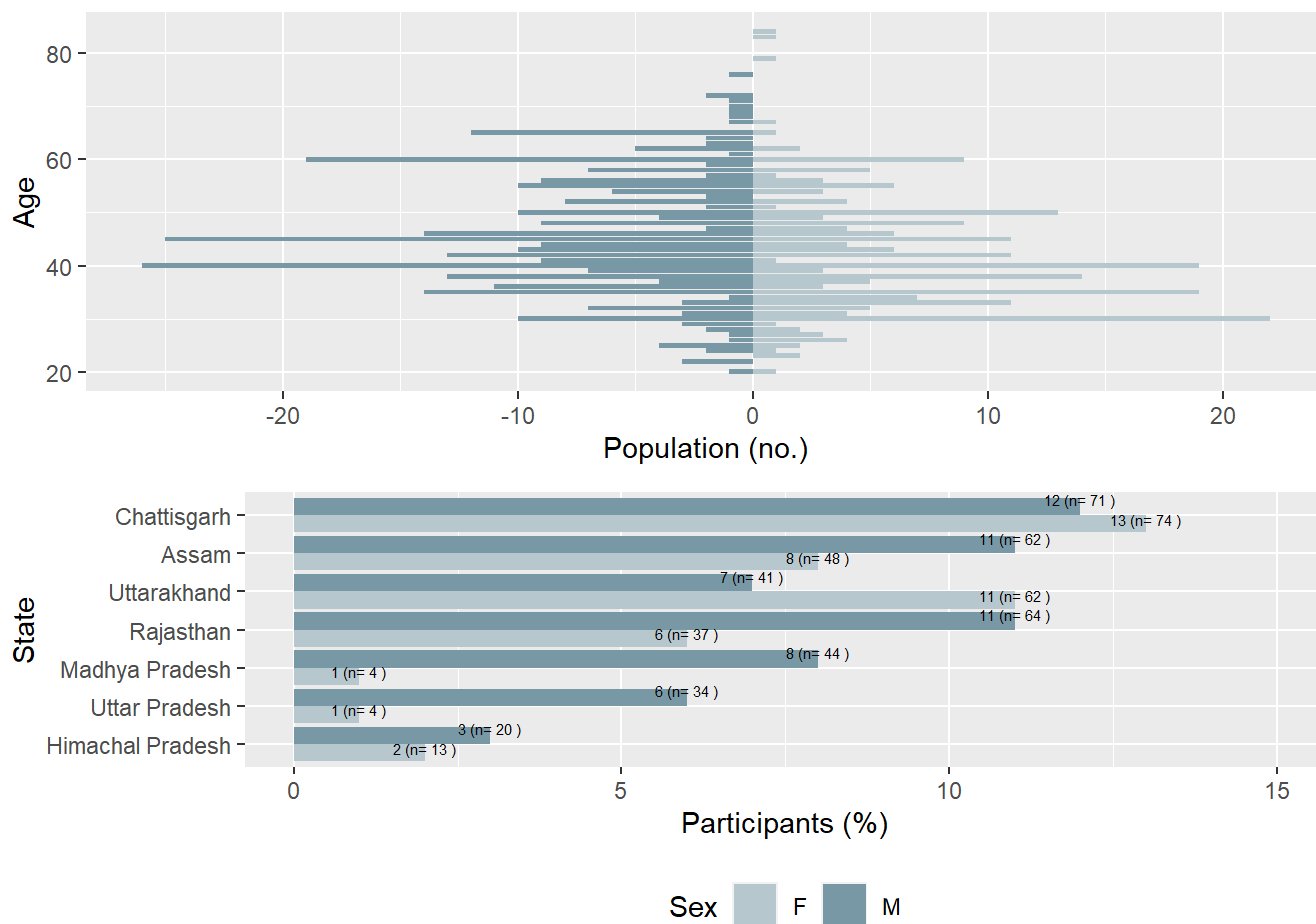
state<-TableB_status_trends%>%
  distinct(Village, State,AEZ)

Participants_state<-Participants%>%
  left_join(state, by="Village")%>%
  group_by(Gender_rev, State)%>%
  summarise(n=n())%>%
  mutate(per=round(n*100/length(Participants$ID),0),
         Sex=if_else(Gender_rev==1, "M", "F"))

state_part<-ggplot(Participants_state,aes(reorder(State, +n),per, fill=Sex))+
  geom_bar(stat = "identity", position = 'dodge')+
  scale_fill_manual(values = c("#B6C7CE", "#7998A5"))+
  theme(legend.position="bottom")+
  ylab("Participants (%)")+
  xlab("State")+
  ylim(0,15)+
  geom_text(aes(label=paste(per, "(n=", n, ")")), position=position_dodge(width=1), v
just=0, size=2)+
  coord_flip()

fig1<-grid.arrange(pop, state_part, nrow=2, ncol=1)

```



```
ggsave("fig1.png",
  plot=fig1,
  width=3.5, height=5, units="in", dpi = 300)
```

####Figure S3 Diversity indexes at the species (species richness) and varietal level (Shannon-Weaver and Pielou's) EH: Eastern Hima-layan; EPH: Eastern Plateau and Hills, WH: Western Himalayan, CPH: Central Plateau and Hills, WDR: Western Dry Region Significant pairwise comparisons Games-Howell test $p \leq 0.05$ are displayed

```

TableB_status_trends<-TableB_status_trends%>%
  left_join(crop_spp_var_names, by=c("Variety name_rev"="Variety name_original2"))

Status_trend<-TableB_status_trends%>%
  mutate(threat=ifelse(Spread_rev==4 | Abundance_rev==4, 5, #lost
    # ifelse(Spread_rev==3 & Abundance_rev==3 | Spread_rev==1 & Abun
    dance_rev==3 | Spread_rev==3 & Abundance_rev==2, 5, #endangered
    ifelse(Spread_rev==1 & Abundance_rev==1, 4, #threatened
    ifelse(Spread_rev==1 & Abundance_rev==2, 3, #nea
r threat
    ifelse(Spread_rev==2 & Abundance_rev==1,
2, #vulnerable
    ifelse(Spread_rev==2 & Abundance_r
ev==2, 0, "nd" )))))) # low concern

AEZ_State<-Status_trend%>%
  group_by(Village, AEZ, State)%>%
  summarise(n=n())%>%
  mutate(AEZ_simple=if_else(AEZ=="Central Plateau and Hills", "CPH",
    if_else(AEZ=="Eastern Himalayan", "EH",
    if_else(AEZ=="Eeastern Plateau and Hills", "EPH",
    if_else(AEZ=="Western Dry Region", "WD
R", "WH" )))))

species_div<-Status_trend%>%
  group_by(botanical_name_rev, Village)%>%
  summarise(n=n())%>%
  spread(botanical_name_rev, n)%>%
  select(-c("Village" , "<NA>" ))

names(species_div)

```

```
## [1] "Amaranthus_hypochondriacus" "Arachis_hypogaea"
## [3] "Brasica_juncia"             "Brassica_rapa"
## [5] "Cajanus_Cajan"              "Cicer_arietinum"
## [7] "Citrullus_colocynthis"      "Citrullus_lanatus"
## [9] "Cuminum_cyminum"            "Cyamopsis_tetragonoloba"
## [11] "Echinochloa_esculenta"      "Eleusine_coracana"
## [13] "Fagopyrum_esculentum"       "Fagopyrum_tataricum"
## [15] "Glycine_max"                 "Guizotia_abyssinica"
## [17] "Hordeum_vulgare"            "Lens_culinaris_"
## [19] "Macrotyloma_uniflorum"      "Oryza_sativa"
## [21] "Panicum_miliaceum"          "Panicum_sumatrense"
## [23] "Paspalum_scrobiculatum"     "Pennisetum_glaucum"
## [25] "Phaseolus_vulgaris"        "Pisum_sativum"
## [27] "Plantago_ovata"             "Ricinus_communis"
## [29] "Sesamum_indicum"            "Setaria_italica"
## [31] "Sorghum_bicolor"            "Trigonella_foenum-graecum"
## [33] "Triticum_aestivum"          "Vigna_aconitifolia"
## [35] "Vigna_mungo"                 "Vigna_radiata"
## [37] "Vigna_umbellata"            "Zea_mays"
```

```
sum(species_div, na.rm = TRUE)
```

```
## [1] 471
```

```
colSums(species_div, na.rm = TRUE)
```

##	Amaranthus_hypochondriacus	Arachis_hypogaea
##	8	2
##	Brasica_juncia	Brassica_rapa
##	15	4
##	Cajanus_Cajan	Cicer_arietinum
##	18	7
##	Citrullus_colocynthis	Citrullus_lanatus
##	1	1
##	Cuminum_cyminum	Cyamopsis_tetragonoloba
##	2	11
##	Echinochloa_esculenta	Eleusine_coracana
##	1	12
##	Fagopyrum_esculentum	Fagopyrum_tataricum
##	2	2
##	Glycine_max	Guizotia_abyssinica
##	5	5
##	Hordeum_vulgare	Lens_culinaris_
##	1	3
##	Macrotyloma_uniflorum	Oryza_sativa
##	2	259
##	Panicum_miliaceum	Panicum_sumatrense
##	1	2
##	Paspalum_scrobiculatum	Pennisetum_glaucum
##	1	9
##	Phaseolus_vulgaris	Pisum_sativum
##	8	2
##	Plantago_ovata	Ricinus_communis
##	1	1
##	Sesamum_indicum	Setaria_italica
##	11	1
##	Sorghum_bicolor	Trigonella_foenum-graecum
##	1	1
##	Triticum_aestivum	Vigna_aconitifolia
##	23	7
##	Vigna_mungo	Vigna_radiata
##	16	15
##	Vigna_umbellata	Zea_mays
##	1	9

```

species_div[is.na(species_div)] <- 0

H <- diversity(species_div[,1:38])
J <- H/log(specnumber(species_div[,1:38]))
spp <- specnumber(species_div[,1:38])

div<-cbind(AEZ_State, H,J, spp )
names(div)<-c("Village","AEZ","State", "n","AEZ_simple" ,"H", "J", "spprich")

H_aez <- ggbetweenstats(
  data = div,
  x = AEZ_simple,
  y = H)+
  labs(
    x = "Agroecological Zone",
    y = "Varietal Shannon-Weaver")+
  theme(
    # Statistical annotations below the main title
    plot.subtitle = element_text(
      family = "arial",
      size = 8,
      face = "bold",
      color="#1b2838"
    )
  )

rich_aez <- ggbetweenstats(
  data = div,
  x = AEZ_simple,
  y = spprich)+
  labs(
    x = "Agroecological Zone",
    y = "Species richness") +
  theme(
    # Statistical annotations below the main title
    plot.subtitle = element_text(
      family = "arial",
      size =8,
      face = "bold",
      color="#1b2838"
    )
  )

j_aez <- ggbetweenstats(
  data = div,
  x = AEZ_simple,
  y = J)+
  labs(
    x = "Agroecological Zone",
    y = "Varietal Pielou's evenness") +

```

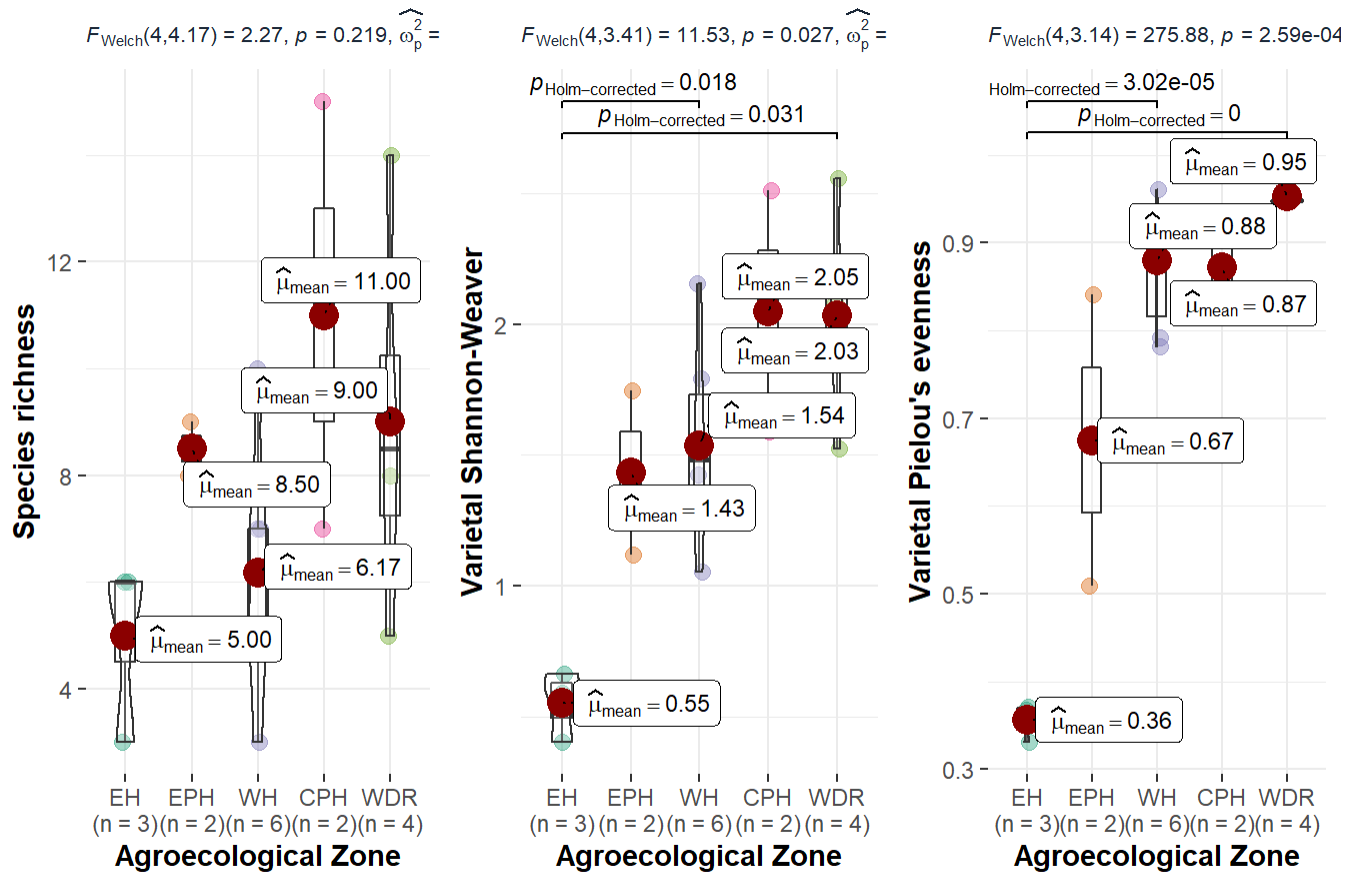


```

theme (
  # Statistical annotations below the main title
  plot.subtitle = element_text(
    family = "arial",
    size = 8,
    face = "bold",
    color="#1b2838"
  )
)

```

```
fig2<-grid.arrange(rich_aez, H_aez, j_aez, nrow=1, ncol=3)
```



est; Comparisons shown: **only significant**

```

ggsave("fig2.png",
  plot=fig2,
  width=12, height=4.5, units="in", dpi = 300)

```

###Figure S4 a) Varieties distribution by threat level, b) major use, c) area change during the last ten years and d) variety type Numbers indicates percentages and number of varieties (in parenthesis) distribution at each threat level – 100% d k data with don't

```

Status_trend_summ<-Status_trend%>%
  filter(!threat=="nd")%>% #Two observations did not include the threat level n=474
  group_by(threat)%>%
  summarise(n=n())%>%
  mutate(percentage=round(n*100/sum(n),0))

threat_cat<-ggplot(Status_trend_summ,aes(x=threat,y=percentage, fill=percentage))+
  geom_bar(stat = "identity")+
  scale_fill_stepsn(breaks = c(25,50), colours = c("#E4F1F6", "#67AFCB"))+
  scale_x_discrete(labels=c( "Least concern","Vulnerable","Near threat", "Threatene
d", "Lost"))+
  theme(legend.position="none")+
  ylab("Varieties (%)")+
  xlab("Theat level")+
  ylim(0,50)+
  geom_text(data=Status_trend_summ,aes(x = threat, y = percentage, label=paste(percen
tage, "(n=", n, ")")),color="black", size=3.5, nudge_x = 0, nudge_y = 5)+
  coord_flip()

Status_trend_use<-Status_trend%>%
  filter(!threat=="nd")%>% #Two observations did not include the threat level n=474
  mutate(major_use_rev1=if_else(major_use_rev=="4"|major_use_rev=="nd", "nd",
                                if_else(major_use_rev=="5"|major_use_rev=="6"|major_u
se_rev=="7"|major_use_rev=="8", "40",major_use_rev )))%>% #5-8=milti-use
  group_by(threat,major_use_rev1)%>%
  summarise(n=n())%>%
  mutate(percentage=round(n*100/sum(n),0))

threat_use<-ggplot(Status_trend_use,aes(x=threat ,y=major_use_rev1, fill=percentag
e))+
  geom_tile()+
  scale_fill_stepsn(n.breaks = 5, colours = c("#E4F1F6", "#67AFCB","#347B98" , "#1A3E
4C", "#092834"))+
  scale_x_discrete(labels=c( "Least concern","Vulnerable","Near threat", "Threatene
d", "Lost"))+
  scale_y_discrete(labels=c( "Sales","Consumption","Fodder", "Multi-uses", "d.k.))+
  theme(legend.position="none",axis.text.y = element_blank())+
  ylab("Major variety use")+
  xlab("")+
  geom_text(data=Status_trend_use,aes(x = threat, y = major_use_rev1, label=paste(per
centage, "(n=", n, ")")),color="white", size=3.5, nudge_x = 0.2, nudge_y = 0)+
  coord_flip()

Status_trend_change<-Status_trend%>%
  filter(!threat=="nd")%>%
  mutate(change_rev1=if_else(change_rev=="4"|change_rev=="nd", "nd",change_rev))%>%
  group_by(threat,change_rev1)%>%
  summarise(n=n())%>%
  mutate(percentage=round(n*100/sum(n),0))

```

```

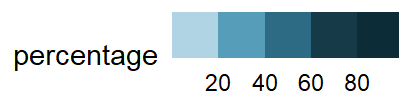
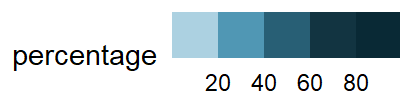
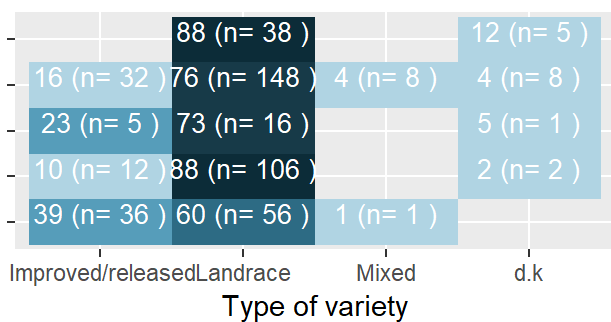
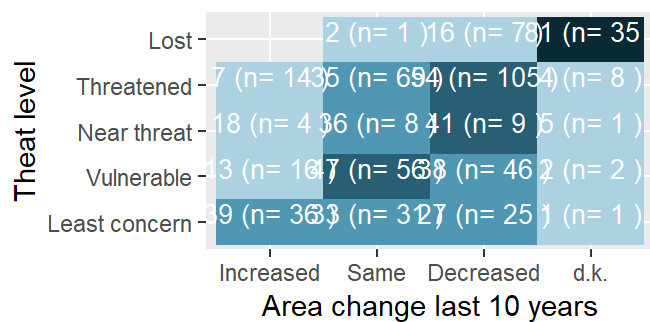
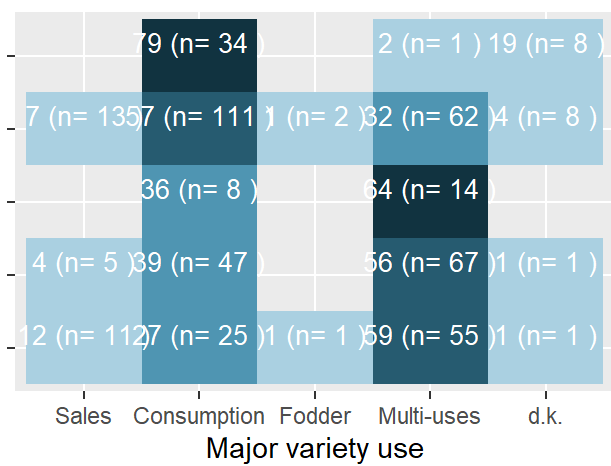
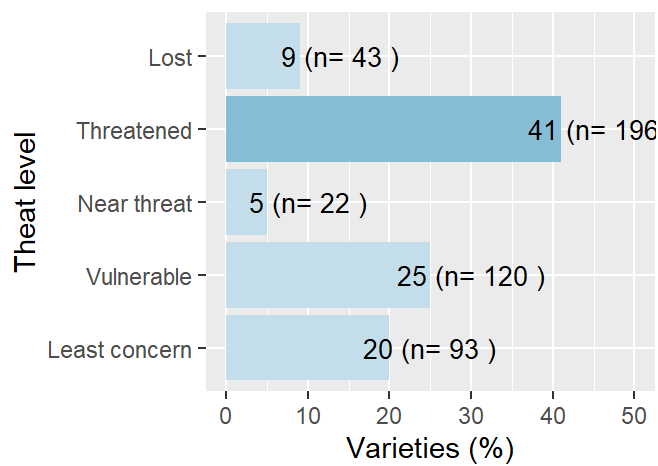
threat_change<-ggplot(Status_trend_change,aes(x=threat ,y=change_rev1, fill=percentage))+
  geom_tile()+
  scale_fill_stepsn(n.breaks = 5, colours = c("#E4F1F6", "#67AFCB","#347B98" , "#1A3E4C", "#092834"))+
  scale_x_discrete(labels=c( "Least concern","Vulnerable","Near threat", "Threatened", "Lost"))+
  scale_y_discrete(labels=c( "Increased","Same","Decreased", "d.k.))+
  theme(legend.position="bottom")+
  ylab("Area change last 10 years")+
  xlab("Theat level")+
  geom_text(data=Status_trend_change,aes(x = threat, y = change_rev1, label=paste(percentage, "(n=", n, ")")),color="white", size=3.5, nudge_x = 0.2, nudge_y = 0)+
  coord_flip()

Status_type<-Status_trend%>%
  filter(!threat=="nd")%>%
  mutate(var_type_rev1=if_else(var_type_rev=="3"|var_type_rev=="nd", "nd",var_type_rev))%>%
  group_by(threat,var_type_rev1)%>%
  summarise(n=n())%>%
  mutate(percentage=round(n*100/sum(n),0))

threat_type<-ggplot(Status_type,aes(x=threat ,y=var_type_rev1, fill=percentage))+
  geom_tile()+
  scale_fill_stepsn(n.breaks = 5, colours = c("#E4F1F6", "#67AFCB","#347B98" , "#1A3E4C", "#092834"))+
  scale_x_discrete(labels=c( "Least concern","Vulnerable","Near threat", "Threatened", "Lost"))+
  scale_y_discrete(labels=c( " Improved/released","Landrace", "Mixed","d.k"))+
  theme(legend.position="bottom" , axis.text.y = element_blank())+
  ylab("Type of variety")+
  xlab("")+
  geom_text(data=Status_type,aes(x = threat, y = var_type_rev1, label=paste(percentage, "(n=", n, ")")),color="white", size=3.5, nudge_x = 0.2, nudge_y = 0)+
  coord_flip()

fig3<-grid.arrange(threat_cat, threat_use, threat_change,threat_type, nrow=2, ncol=2)

```



```
ggsave("fig3.png",
  plot=fig3,
  width=9, height=8, units="in", dpi = 300)
```

####Figure S5 Varietal risk index at the agroecological and site level and by variety type Numbers indicate the number of varieties

```

var_vill_state_aez<-Status_trend%>%
  distinct(Village, State, AEZ)%>%
  mutate(AEZ_simple=if_else(AEZ=="Central Plateau and Hills", "CPH",
                            if_else(AEZ=="Eastern Himalayan", "EH",
                                    if_else(AEZ=="Eeastern Plateau and Hills", "EPH",
                                            if_else(AEZ=="Western Dry Region", "WD
R", "WH" )))))

Status_trend_var<-Status_trend%>%
  filter(!threat=="nd")%>%
  filter(var_type_rev=="1"|var_type_rev=="2")%>%
  group_by(Village, threat, var_type_rev )%>%
  summarise(n=n(), max_threat=max(threat))%>%
  mutate(Weighted_score=n*as.numeric(threat))

Status_trend_var_tot<-Status_trend_var%>%
  group_by(Village)%>%
  summarise(max_score=max(threat), n_tot=sum(n))%>%
  mutate(max=n_tot*as.numeric(max_score))%>%
  select(-c(max_score))

Status_trend_var<-Status_trend_var%>%
  left_join(Status_trend_var_tot, by="Village")

Status_trend_Score<-Status_trend_var%>%
  group_by(Village, var_type_rev, n_tot, max)%>%
  summarise(score_t2=sum(Weighted_score), max_score=max(threat), no_var=sum(n))%>%
  mutate(var_threat_index=(1-(max-score_t2)/(max)),
         type=if_else(var_type_rev=="1", "Improved", "Landrace"))%>%
  left_join(var_vill_state_aez, by="Village")

Status_trend_Score_mean<-Status_trend_Score%>%
  group_by(AEZ, type )%>%
  summarise(mean=mean(var_threat_index))

mean_landrace<-Status_trend_Score_mean%>%filter(type=="Landrace")
mean_improved<-Status_trend_Score_mean%>%filter(type=="Improved")

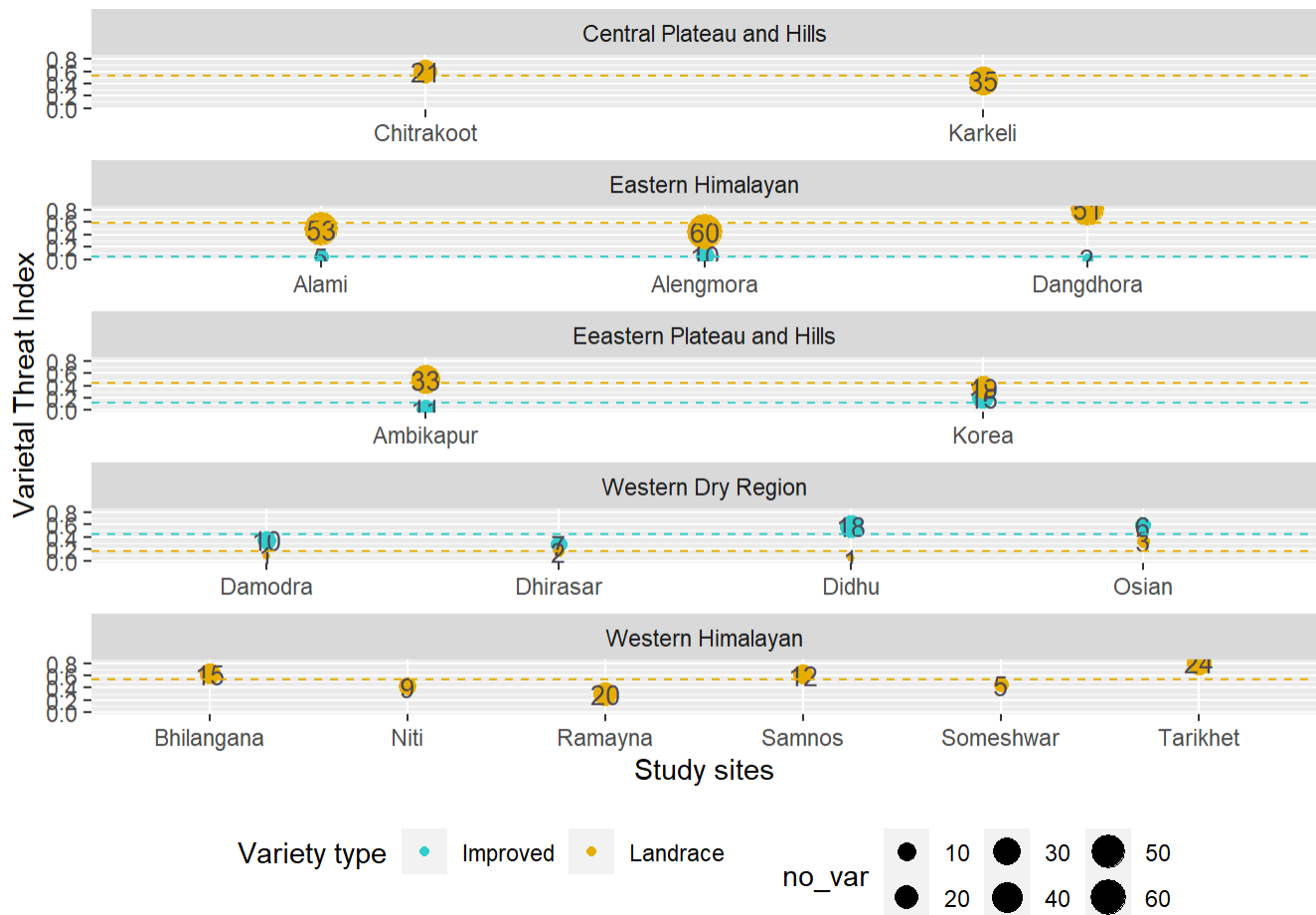
fig4<-ggplot(data=Status_trend_Score, aes_string(x="Village", y="var_threat_index" , co
lour="type", size="no_var"))+
  geom_point()+
  scale_colour_manual(name="Variety type", values = c("#33cccc", "#e6ad00"))+
  #labs(shape="No. varieties", colour="Variety type")+
  geom_text(data=Status_trend_Score, aes(x = Village, y = var_threat_index, label=no_v
ar), color="#494452", size=3.5)+
  ylab("Varietal Threat Index")+
  xlab("Study sites")+
  facet_wrap(~AEZ, ncol=1, strip.position = "top", scales = "free_x")+

```

```

geom_hline(data= mean_landrace, aes(yintercept=mean), color="#e6ad00", linetype="dashed")+
  geom_hline(data= mean_improved, aes(yintercept=mean), color="#33cccc", linetype="dashed")+
  theme(legend.direction = "horizontal", legend.position = "bottom", legend.box = "horizontal")
fig4

```



```

ggsave("Fig4.png",
  plot=fig4,
  width=6, height=6.5, units="in", dpi = 300)

```

####Figure S6 Sanki diagram showing the varietal flow and distribution between the type (landrace or improved/released), across food groups and agroecological zones. Each food group includes the number of species-spp and varieties-var listed. Food groups include Grains (e.g., wheat, barley, millets, maize, sorghum, rice), Pulses (e.g., gram, lentils, beans, peas), Cond & DGLV - Condiments and seasonings and Dark green leafy vegetables (e.g., mustard, fenugreek), Grains & DGLV – Grains and Dark green leafy vegetables (e.g., amaranth), Nuts and seeds (e.g., sesame, groundnut), Oil (e.g., niger), Medicinal (e.g., Ricinus, Isabgol), Other fruits (e.g., watermelon) and Other veg – vegetables (e.g., cucumber).

