

# Compare All

Ariella Fuzaylov

3/29/2021

## Loading Data From CSV and Update Header

```
my.data=read.csv("Online Recipe Sharing.csv", header=TRUE)
colnames(my.data)<-c("Timestamp", "Age", "Primary.Meal.Prepper", "Household.Dietary.Restriction",
"Home.Cooking.Rate",
"Primary.Recipe.Format",
"Primary.Search.Website",
"Enjoyed.Website.Searching", "Comments.Enjoyed.Website.Searching", "NOT.Enjoyed.Website.Searching", "Comments.Enjoyed.Website.Browsing", "Comments.NOT.Enjoyed.Website.Browsing",
"Previous.Recipe.Search.Frequency",
"Browsing.While.Searching.Frequency",
"Click.Rate",
"Search.Browse.Same.Websites",
"Primary.Browsing.Website",
"Enjoyed.Website.Browsing",
"Comments.Enjoyed.Website.Browsing", "NOT.Enjoyed.Website.Browsing", "Comments.NOT.Enjoyed.Website.Browsing",
"Source.of.Influential.Reviews", "Frequency.Reviews.Effect.Behavior",
"Frequency.Seek.Out.Review",
"Frequency.of.Review",
"Frequency.of.Recipe.Saving",
"Method.of.Recipe.Saving",
"Modification.Frequency",
"Modification.Influence.Factors",
"Modification.Record.Frequency",
"Modification.Record.Method",
"Satisfaction.with.Available.Record.Methods",
"Interest.in.Improved.Record.Method",
"Frequency.of.Recipe.Discussion", "Frequency.of.Reading.Discussion",
"Primary.Discussion.Medium", "Enjoyed.Features.of.Discussion.Mediums", "Ingredients.L.V.Above",
"Ingredients.L.Comments.Inline.V.Below", "Ingredients.Above.Comments.Below.V.Inline", "Ingredients.By.Step.V.Scroll.L",
"Ingredients.By.Step.V.Scroll.L",
"Ingredients.Above.V.Scroll.L")
```

## Re-Factor Data

If Respondent indicated that they search and browse on the same websites, populate the empty cells with the same data. This assumes that the user's searching behavior is exactly the same as the browsing behavior if the user selected yes for searching and browsing on the same websites.

```

for (i in 1:nrow(my.data)){
  if (my.data$Search.Browse.Same.Websites[i]=="No"){
    my.data$Primary.Browsing.Website[i]<-my.data$Primary.Search.Website[i]
    my.data$Enjoyed.Website.Browsing[i]<-my.data$Enjoyed.Website.Searching[i]
    my.data$NOT.Enjoyed.Website.Browsing[i]<-my.data$NOT.Enjoyed.Website.Searching[i]
  }
}

```

Since the data set is small, I am consolidating some of the categories.

- Primary Meal Prepper will be Respondent if the individual taking the survey indicated that they are the primary meal prepper in their household or if they cook for themselves, and other in all other cases.
- Dietary restriction will become a yes or no question
- Home Cooking Rate will become Daily if the respondents cooks at home most days, weekly if the respondent cooks several times a week, and monthly is the respondent cooks a couple times a month.

```

my.data.factor<-my.data
my.data.factor$Age<-as.factor(my.data$Age)

my.data.factor$Primary.Meal.Prepper<-as.factor(my.data.factor$Primary.Meal.Prepper)

my.data.factor$Household.Dietary.Restriction<-as.factor(my.data.factor$Household.Dietary.Restriction)

my.data.factor$Home.Cooking.Rate<-as.factor(my.data.factor$Home.Cooking.Rate)

my.data.factor$Ingredients.L.V.Above<-as.factor(my.data.factor$Ingredients.L.V.Above)
my.data.factor$Ingredients.By.Step.V.Above<-as.factor(my.data.factor$Ingredients.By.Step.V.Above)
my.data.factor$Ingredients.Above.V.Scroll.L<-as.factor(my.data.factor$Ingredients.Above.V.Scroll.L)
my.data.factor$Ingredients.L.Comments.Inline.V.Below<-as.factor(my.data.factor$Ingredients.L.Comments.Inline.V.Below)
my.data.factor$Ingredients.By.Step.V.Scroll.L<-
  as.factor(my.data.factor$Ingredients.By.Step.V.Scroll.L)
my.data.factor$Ingredients.Above.Comments.Below.V.Inline<-
  as.factor(my.data.factor$Ingredients.Above.Comments.Below.V.Inline)
my.data.factor<- mutate(my.data.factor,
  Age = fct_collapse(Age,
    YA = c("18 - 24 years old", "25 - 34 years old"),
    Adult = c("35 - 44 years old", "45 - 54 years old", "55 - 64 years old"),
    other_level = "Other"),
  Primary.Meal.Prepper = fct_collapse(Primary.Meal.Prepper,
    Respondent = c("You", "I cook for myself"),
    other_level = "Other"),
  Household.Dietary.Restriction=fct_collapse(Household.Dietary.Restriction,
    No="None",
    other_level = "Yes"),
  Home.Cooking.Rate=fct_collapse(Home.Cooking.Rate,
    Daily=c("Almost every meal", "Daily", "Every meal"),
    Weekly=c("Several times a week", "Once or twice a week"),
    Monthly=c("Once or twice a month")),
  Ingredients.L.V.Above=fct_collapse(Ingredients.L.V.Above,
    Ing.L =c("A"),
    Ing.Above=("B")),
  Ingredients.By.Step.V.Above=fct_collapse(Ingredients.By.Step.V.Above,
    Ing.By.Step=c("A"),

```

```

Ing.Abov=c("B")),
Ingredients.Above.V.Scroll.L=fct_collapse(Ingredients.Above.V.Scroll.L,
Ing.Above=c("A"),
Scroll.L=c("B")),
Ingredients.L.Comments.Inline.V.Below=fct_collapse(Ingredients.L.Comments.Inline.V.Below,
Ing.L.Com.Inline=c("A"),
Ing.L.Com.Below=c("B")),
Ingredients.By.Step.V.Scroll.L=fct_collapse(Ingredients.By.Step.V.Scroll.L,
Ing.By.Step=c("A"),
Ing.Scroll=c("B")),
Ingredients.Above.Comments.Below.V.Inline=fct_collapse(Ingredients.Above.Comments.Below.V.Inline,
Ing.Above.C.Below=c("A"),
Ing.Above.C.Inline=c("B"))
)

```

### Website Recoding:

For the sake of this analysis any website that has a test kitchen that creates editorial content or is able to curate content from professional sources is a magazine, a website with one or two people testing recipes is a blog, and a website that allows users to contribute their own recipes is community based. The information for this classification is found on the website's about page. Additionally, media such as cookbooks and podcasts are classified under Influencers due to their personality driven nature.

### Discussion Method Recoding:

Any type of online chatting be it texting, discord, etc. has been grouped together into Digital Chat. Any type of interpersonal communication where a chat method was not specified is grouped into verbal.

Note saving methods that mention remembering or memory are grouped into memory, while respondents that indicate that they do not take any type of notes and do not try to remember are grouped into None.

### Modification Recoding:

Modification influence factors pertaining to diet, or nutrition are grouped together under the umbrella of "Diet".

Modification influence factors pertaining to personal preference for food, flavor, or preparation method are grouped together under the category of "Personal Preference".

Modification influence factors pertaining ingredients availability are grouped together under the category of "Ing. Availability"

```
unique(separate_rows(my.data.factorred[32],1, sep = ";"))
```

```

## # A tibble: 18 x 1
##   Modification.Record.Method
##   <chr>
## 1 ""
## 2 "None"
## 3 "Mentally?"
## 4 "Digital notes"

```

```
## 5 "Physical notes"
## 6 "Mental note"
## 7 "I don't :o"
## 8 "Comments section provided for recipe"
## 9 "Memory"
## 10 "N/A"
## 11 "I mostly just remember it for next time "
## 12 "brainpower"
## 13 "I dont"
## 14 "I don't"
## 15 "I store it in my noggin"
## 16 "I don't.."
## 17 "i dont"
## 18 "i don't"
```

```
my.data.selected<-my.data.factoried[c(6,7,8,10,17,18,20,22,23,28,37,30,32,38)]
variables<-c()
```

*##This creates a vector that will recode the variables with the proper names*

```
for (i in 1:ncol(my.data.selected)){
  temp<- my.data.selected[i]
  temp<-separate_rows(temp,1, sep = ";")
  variables<-append(variables,temp[[1]])
  variables<-unique(variables)
  data.frame(variables)
}
```

```
cleaned.variables<-c(
```

```
"Mobile",
"Desktop",
"Digital",
"Physical Print",
"Physical Print",
"Digital",
"Physical Family",
"Physical Family",
"Physical Family",
"Physical Family",
"Mags",
"Blogs",
"Google",
"Video",
"Community Based" ,
"Mags",
"Community Based" ,
"Pinterest",
"Blogs",
"Video",
"Mags",
"Facebook",
"Reddit",
"Mags",
"Mags",
"Mags",
```



```

"Diet",
"Diet",
"Preference",
"Diet",
"Ing. Availability",
"Ing. Availability",
"Recommendation",
"Preference",
"Memory",
"Digital",
"Physical",
"Memory",
"None",
"Comments",
"None",
"Memory",
"Memory",
"None",
"None",
"Memory",
"None",
"None",
"5 Star Review",
"Groups",
"Up/Down Vote Posts",
"Up/Down Vote Com.",
"Collapse Comment",
"Comment Reply",
"Comment Thread",
"Inline Comment")

names(cleaned.variables)<-variables

```

## Functions for Cleaning Data

```

dummies<-function(search.data, to.clean){
  col.names<-c(names(search.data))
  col.names<-col.names[col.names!=to.clean]
  search.data.clean<- search.data%>% separate_rows(all_of(to.clean), sep = ";")

  search.data.clean[to.clean]<-
    as.character(cleaned.variables[search.data.clean[[to.clean]])]
  search.data.clean[to.clean]<-lapply(search.data.clean[to.clean],function(x) replace(x,is.na(x),"Empty"))

  search.data.dummies<-search.data.clean%>%
    select((to.clean))%>%
    dummy()%>%
    bind_cols(search.data.clean)%>%
    select(-(to.clean))%>%
    pivot_longer(cols=-col.names, names_to = "key", values_to = "value")%>%
    filter(value!=0)

```

```

search.data.dummies<-search.data.dummies%>%
  unique()

search.data.dummies<-search.data.dummies%>%
  spread(key, value, fill = 0)

}

```

## Load Factored Data

```

search.data<-my.data.factored[-c(1,9,11,19,21)]
search.data<-data.frame(search.data)
new.names=c("Age", "Meal.Prepper", "Dietary.Restriction", "Home.Cook.Rate", "Primary.Format.C", "Primary.S.",
  "Enjoyed.S.C", "NOT.Enjoyed.S.C", "Recipe.Search.F", "Repeat.S.F", "Browse.Search.F", "Click.Rate",
  "Search.Browse.Same", "Primary.B.C", "Enjoyed.B.C", "NOT.Enjoyed.B.C", "Primary.R.C", "Influenced",
  "Use.R.F", "Seek.R.F", "R.F", "Save.F", "Save.C", "Mod.F", "Why.Mod.C", "Mod.Note.F", "Mod.Note.C",
  "Note.Method.S", "Potential.Note.Taker", "Disc.F", "Read.Disc.F", "Disc.C", "Enjoy.Disc.C", "Ing.L.Com.Inline.V.Below",
  "Ing.Above.Com.Below.V.Inline", "Ing.By.Step.V.Above", "Ing.By.Step.V.Below", "Ing.Above.V.Scroll.L"
)
colnames(search.data)<-new.names

search.data<-tibble::rowid_to_column(search.data, "ID")

# for (col in colnames(select(search.data,ends_with(".S")))){
#   search.data[[col]]<-factor(search.data[[col]], levels=c(NA,"1","2","3","4","5"))
#   levels(search.data[[col]])<- c("Dissatisfied","Somewhat Dissatisfied", "Neutral",
#     "Somewhat Satisfied","Satisfied")
# }
#
# for (col in colnames(select(search.data,ends_with(".F")))){
#   search.data[[col]]<-factor(search.data[[col]], levels=c(NA,"1","2","3","4","5"))
#   levels(search.data[[col]])<- c("Never","Rarely","Sometimes", "Often","Always")
# }

```

The categories that are farther from the origin explain more of the variance in the data set, and are well represented by the factor map. Furthermore, the categories that are closer together have similar profiles.

From this data we can see that frequency of discussion, seeking reviews, reading discussion, reviewing, and saving recipes have similar profiles. Additionally, neither the first nor the second component have negatively correlated variable categories.

## Improved MCA

```

cleaned<-search.data
to.dummy<-select(cleaned, ends_with(".C"))
to.dummy.cols<-c(colnames(to.dummy))

for (col in to.dummy.cols){
  cleaned<-dummies(cleaned,c(col))
}

```

```
}
```

```
## Note: Using an external vector in selections is ambiguous.  
## i Use 'all_of(to.clean)' instead of 'to.clean' to silence this message.  
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.  
## This message is displayed once per session.
```

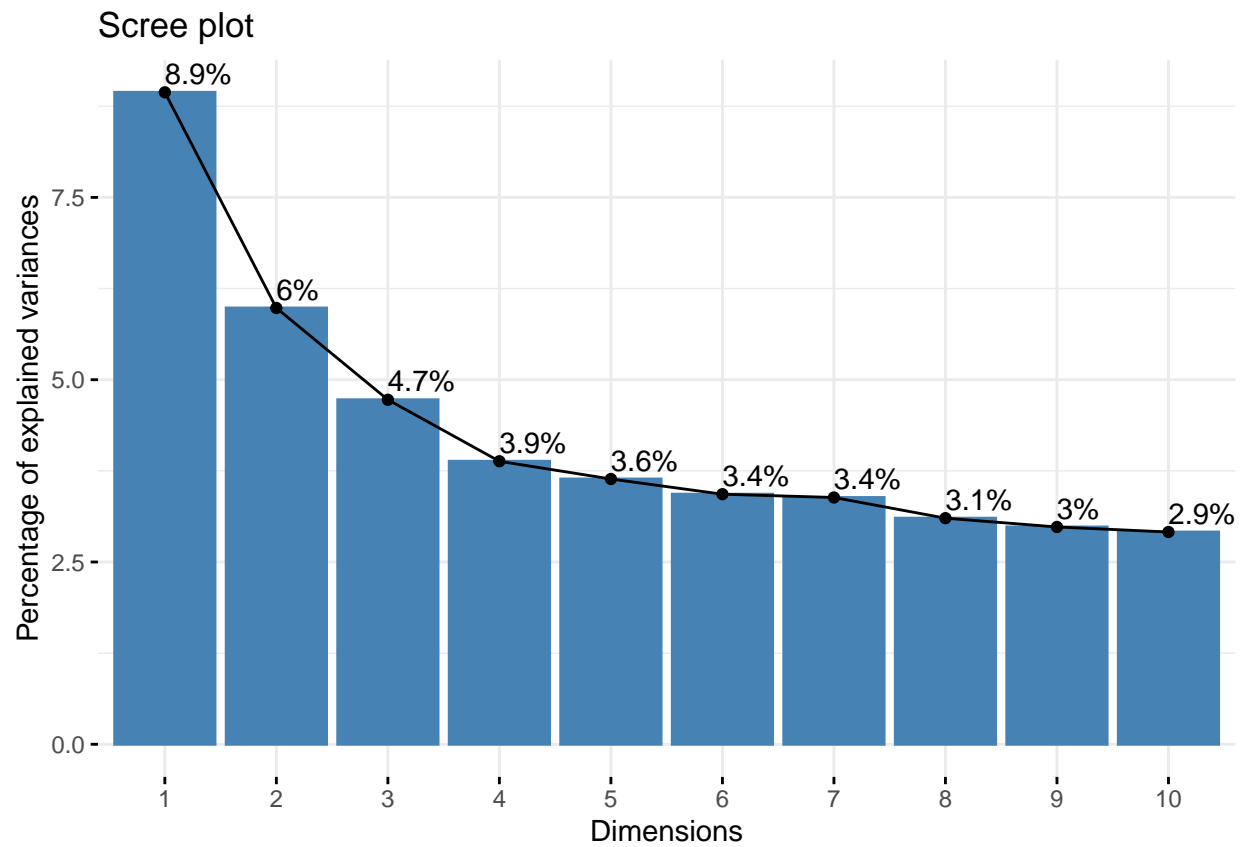
```
## Note: Using an external vector in selections is ambiguous.  
## i Use 'all_of(col.names)' instead of 'col.names' to silence this message.  
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.  
## This message is displayed once per session.
```

```
cols<-names(cleaned)  
cleaned.factored<-lapply(cleaned[cols], as.factor)
```

```
cleaner.S<-function(df){  
  to.dummy<-select(df, ends_with(".C"))  
  to.dummy.cols<-c(colnames(to.dummy))  
  
  for (col in to.dummy.cols){  
    df<-dummies(df,c(col))  
  }  
  
  cols<-names(df)  
  cleaned.factored<-lapply(df[cols], as.factor)  
  cleaned.table<-data.frame(cleaned.factored[-c(1)])  
}
```

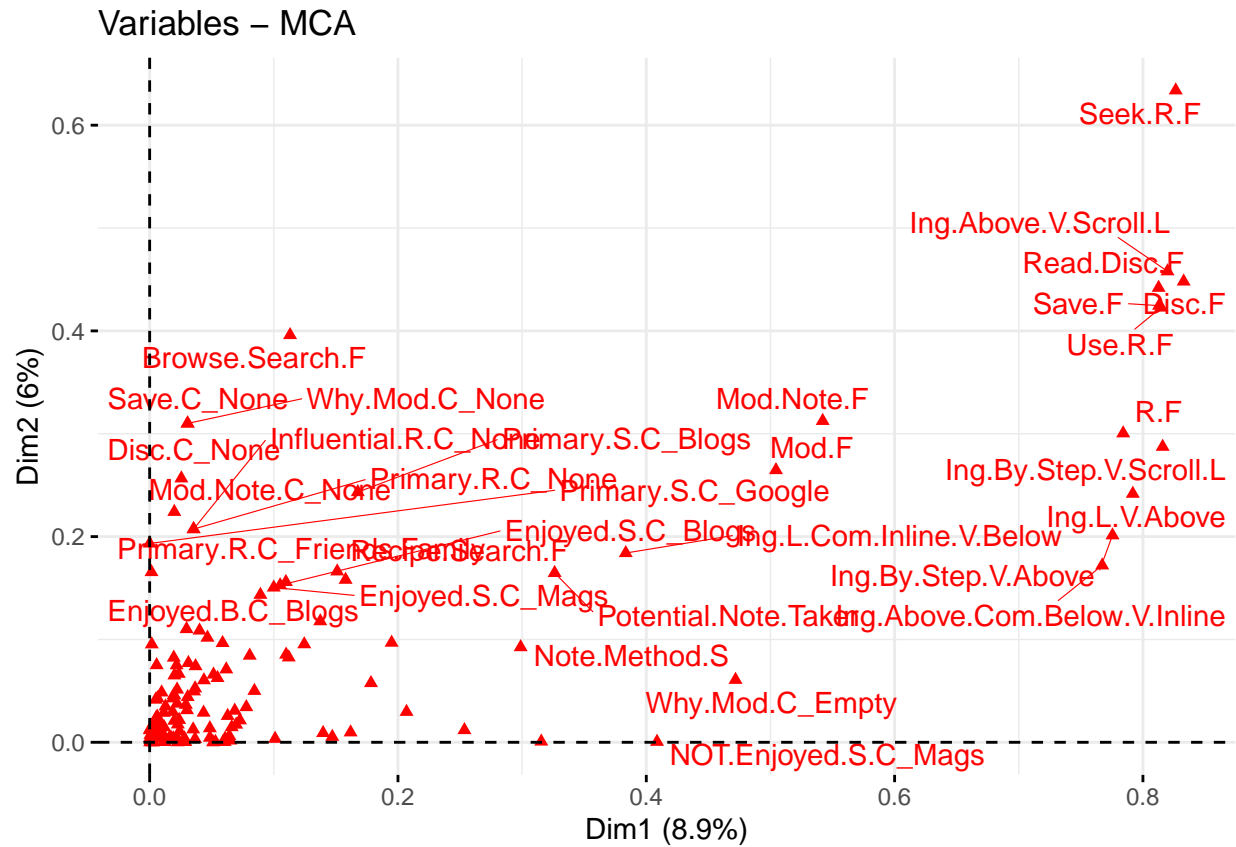
```
cleaned.search.data<-data.frame(cleaned.factored[-c(1)])  
search.MCA=MCA(cleaned.search.data,graph=FALSE)  
fviz_screplot(search.MCA,addlabels=T)
```





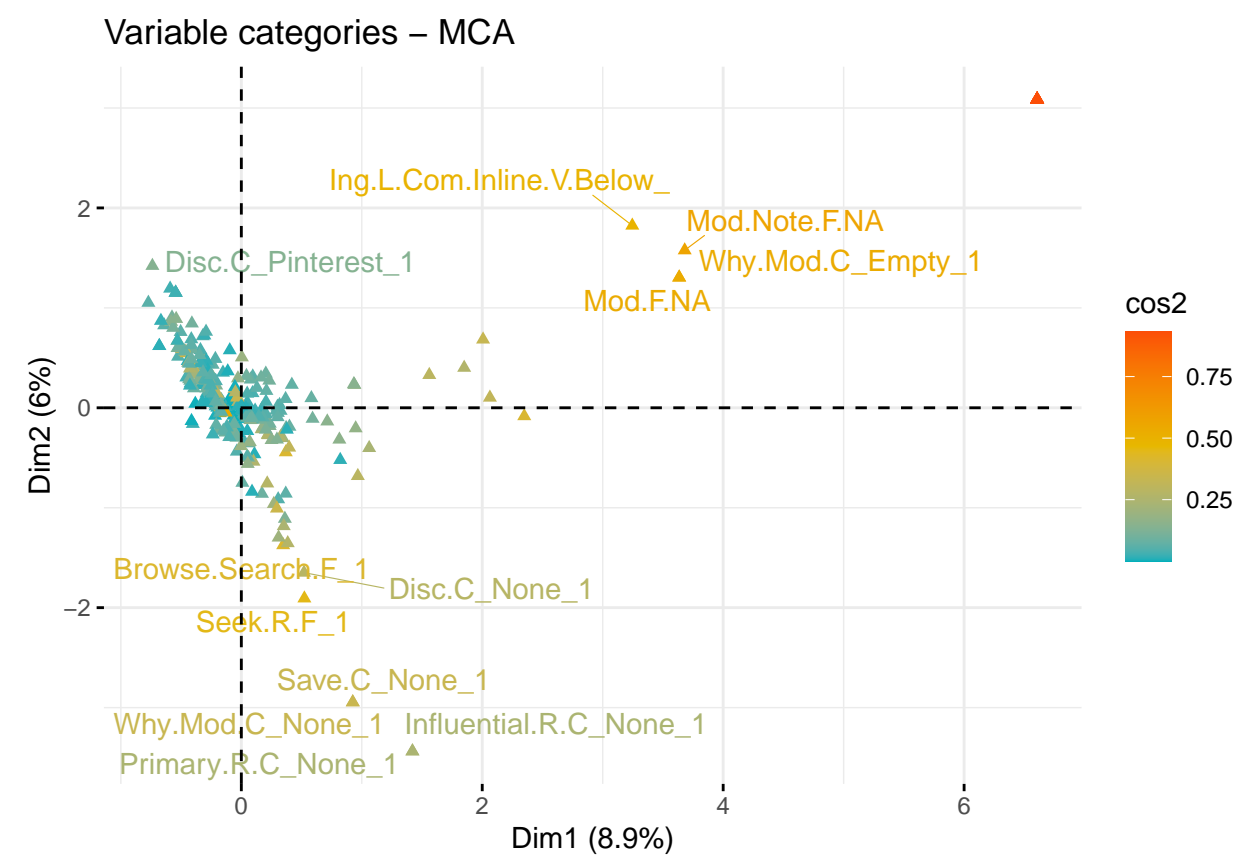
```
fviz_mca_var(search.MCA, choice = "mca.cor", repel = TRUE,  
             ggtheme = theme_minimal())
```

```
## Warning: ggrepel: 110 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```



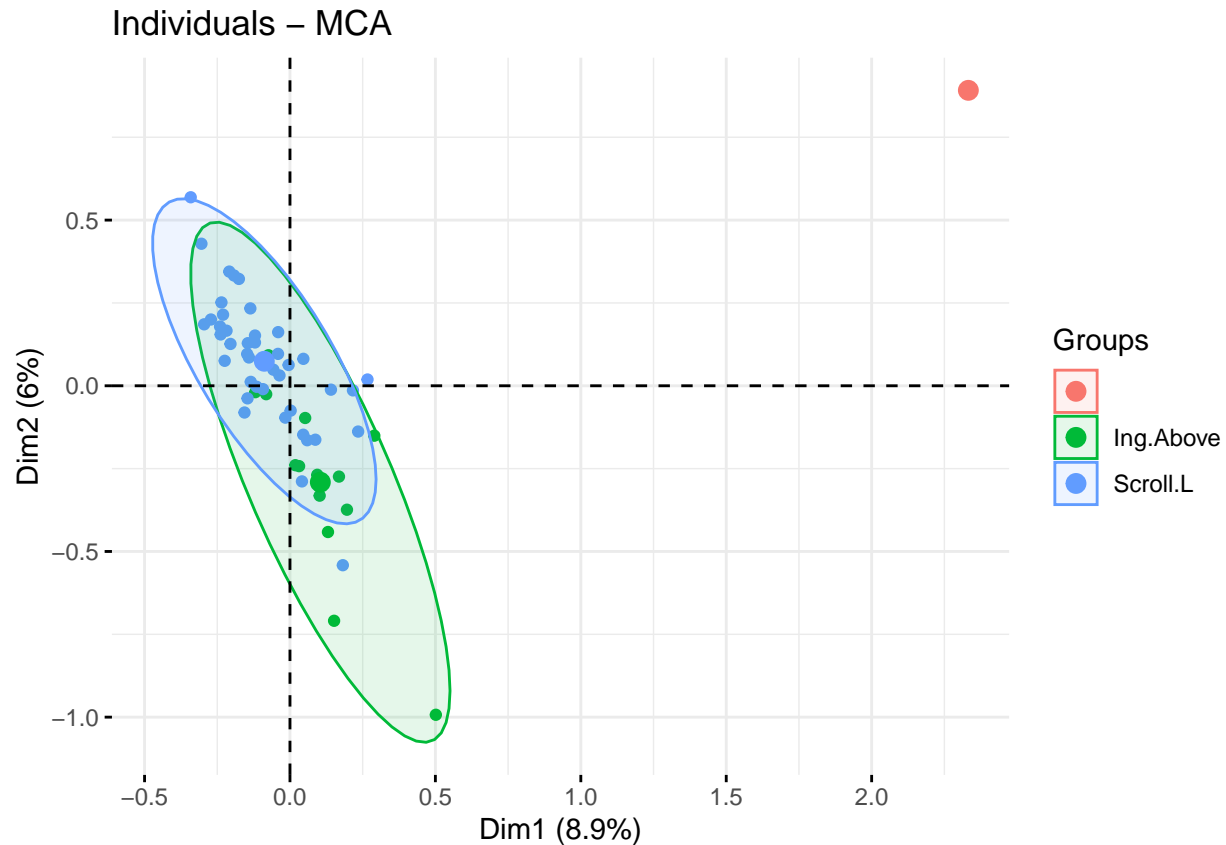
```
fviz_mca_var(search.MCA, col.var = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE, ggtheme = theme_minimal())
```

```
## Warning: ggrepel: 328 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
fviz_mca_ind(search.MCA, label="none", habillage = cleaned.search.data$Ing.Above.V.Scroll.L, addEllipse
```

```
## Too few points to calculate an ellipse
```



#### Scree Plot

Using the elbow rule, I will be looking at the first two components of this MCA analysis. Therefore this plot will only capture 14.6% of the variability in the data.

**Variables** Seeking Reviews is well represented and correlated with both dimensions.

Similar Profiles:

- No saving method, no modification note method, and no primary recipe website
- Not enjoying recipe magazines & Empty why modify answer
- Reading Discussion Frequency, Discussion Frequency, Use reviews frequency, Save frequency, AB: Ingredients above vs. scroll left
- Modification Frequency, Modification Note Frequency, Ingredient listed step by step vs. ingredients pinned left when scrolling

**Variable Categories** Users that did not answer questions are associated in this analysis.

```
mca.plot<-function(selected.data){
  selected.data.cleaned<-cleaner.S(selected.data)
  search.MCA=MCA(selected.data.cleaned,graph=FALSE)
  scree<-fviz_screplot(search.MCA,addlabels=T)
  contrib<-fviz_mca_var(search.MCA,
                        choice = "mca.cor", repel = TRUE,ggtheme = theme_minimal())

  biplot<-fviz_mca_var(search.MCA, col.var = "cos2",
```

```

        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
        repel = TRUE, ggtheme = theme_minimal())
print(screes)
print(contrib)
print(biplot)
return(search.MCA)
}

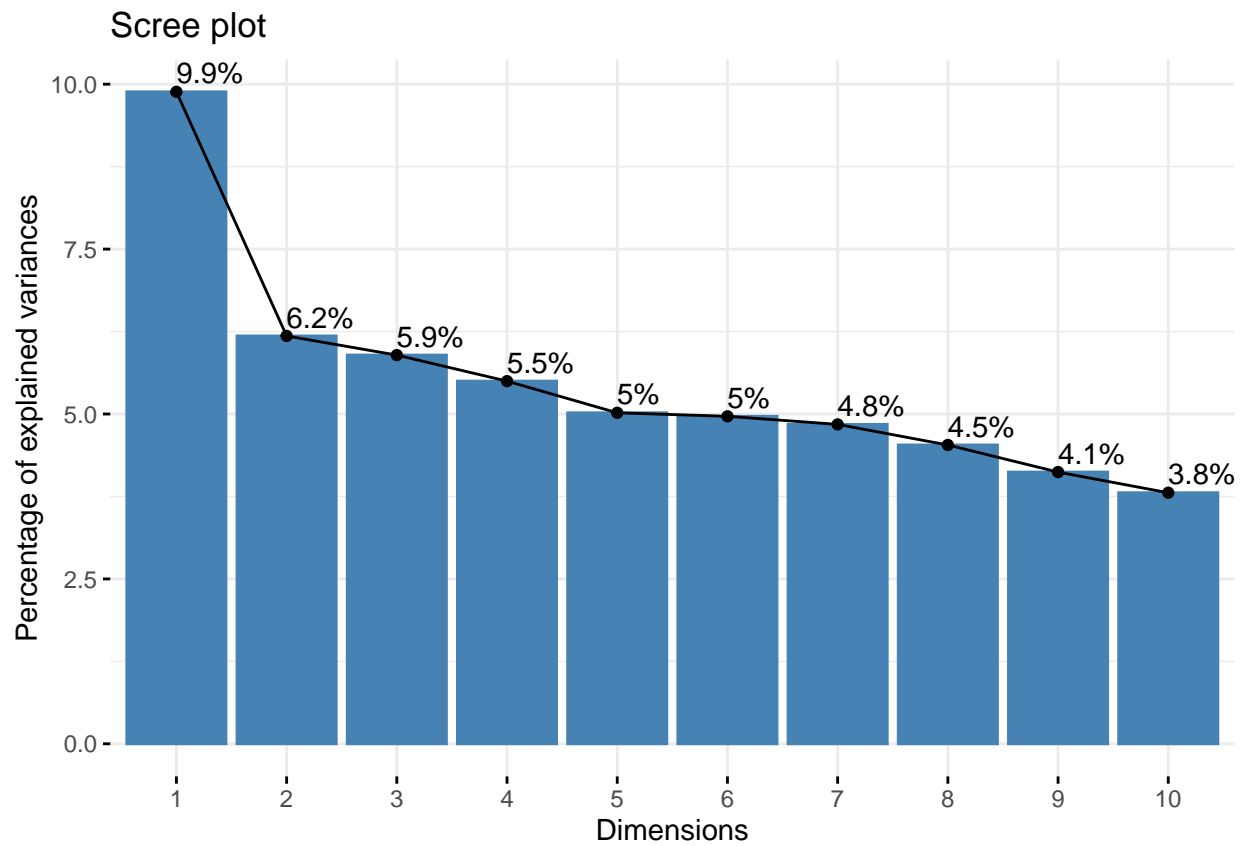
```

## Stratify By Diet

```

cleaned.Diet.Yes<-filter(search.data, Dietary.Restriction == "Yes")
cleaned.Diet.Yes<-cleaned.Diet.Yes%>%select(-c(Dietary.Restriction))
cleaned.data<-mca.plot(cleaned.Diet.Yes)

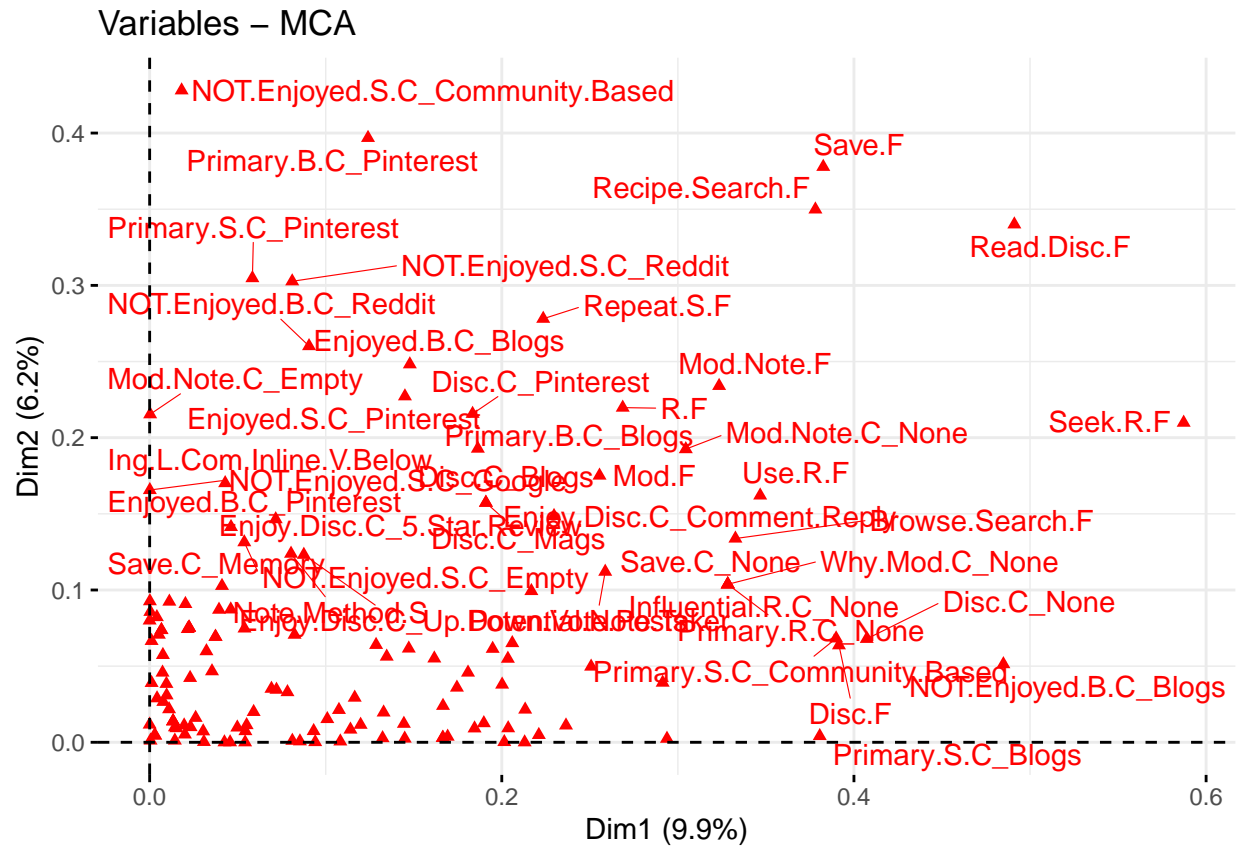
```



```

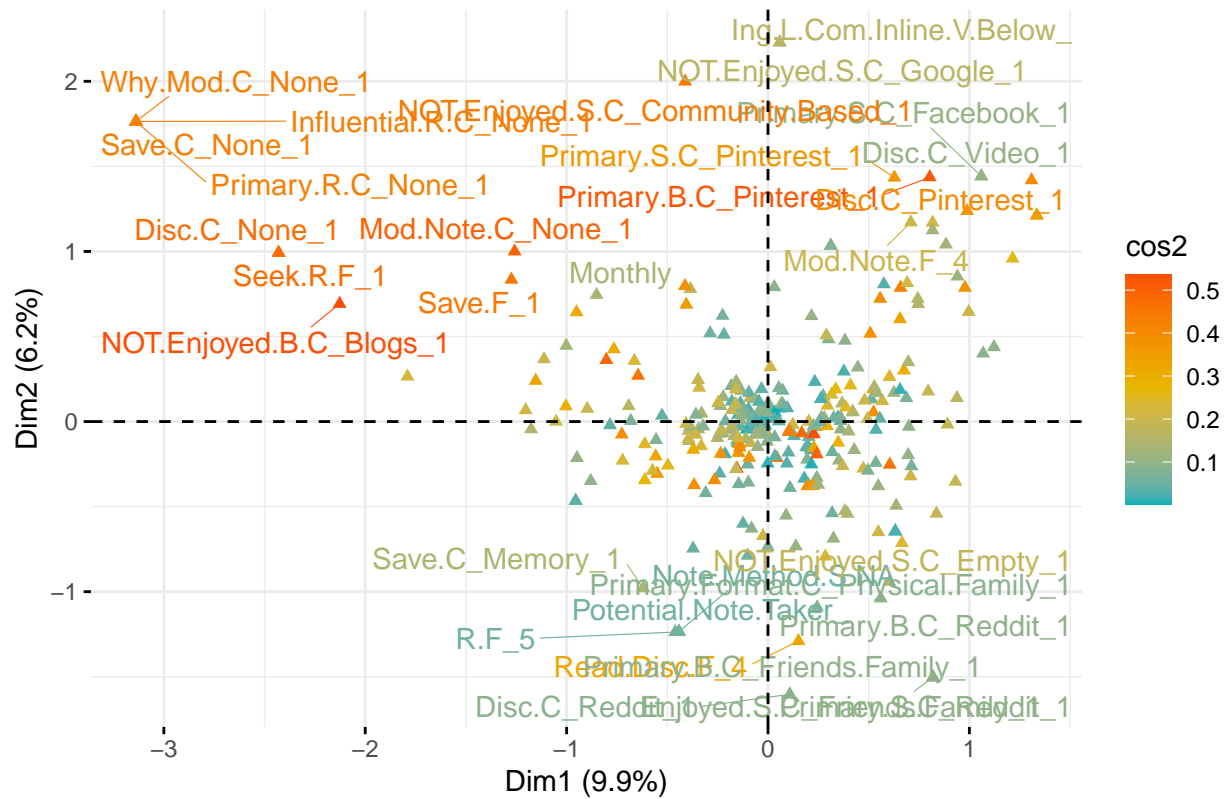
## Warning: ggrepel: 97 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```

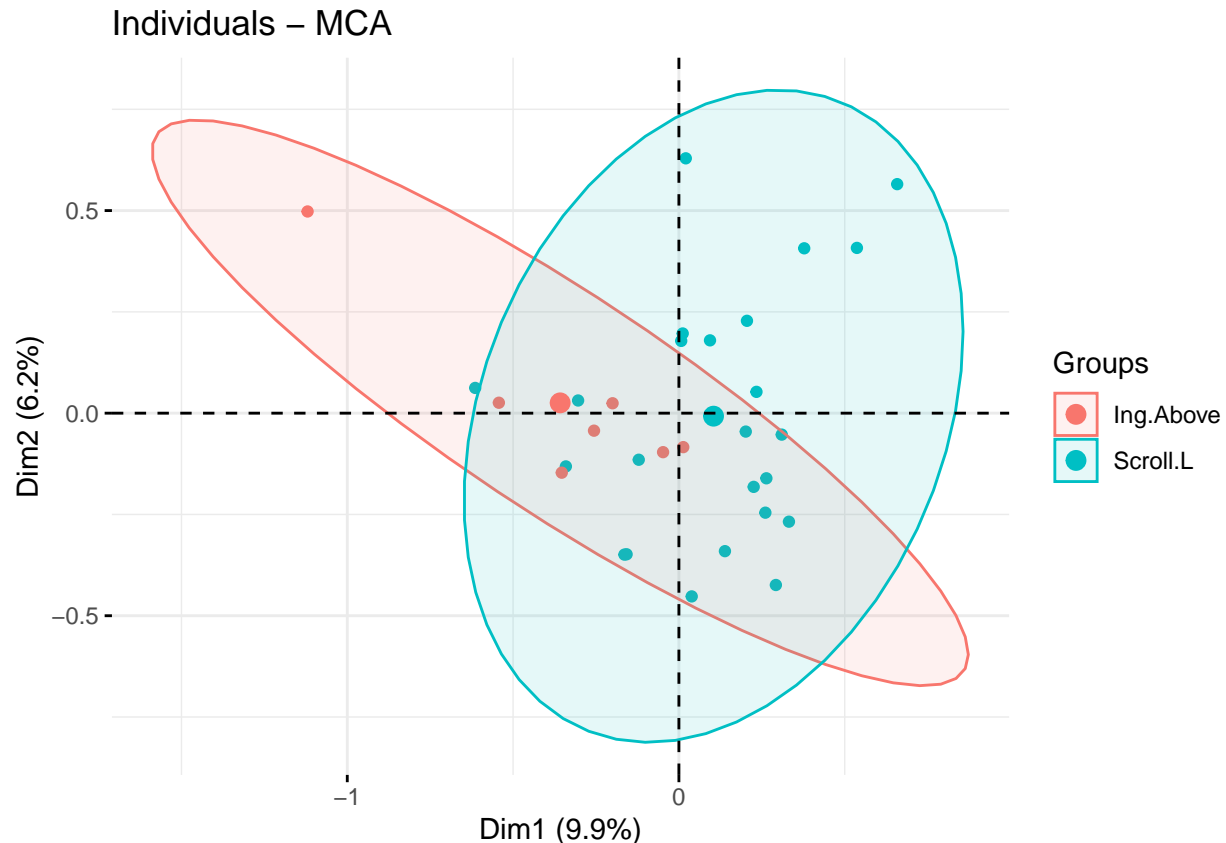


```
## Warning: ggrepel: 287 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

## Variable categories – MCA



```
fviz_mca_ind(cleaned.data, label="none", habillage = cleaned.Diet.Yes$Ing.Above.V.Scroll.L, addEllipses
```



#### #### Scree Plot

Using the elbow rule, I will be looking at the first two components of this MCA analysis. Therefore this plot will only capture 16.2% of the variability in the data. Slightly More variance is captured in the components when we stratify by diet.

**Variables** Save frequency, Recipe search frequency, and Reading the recipe discussion frequency have similar profiles in individuals who do follow a diet. These three columns are well represented by both dimensions.

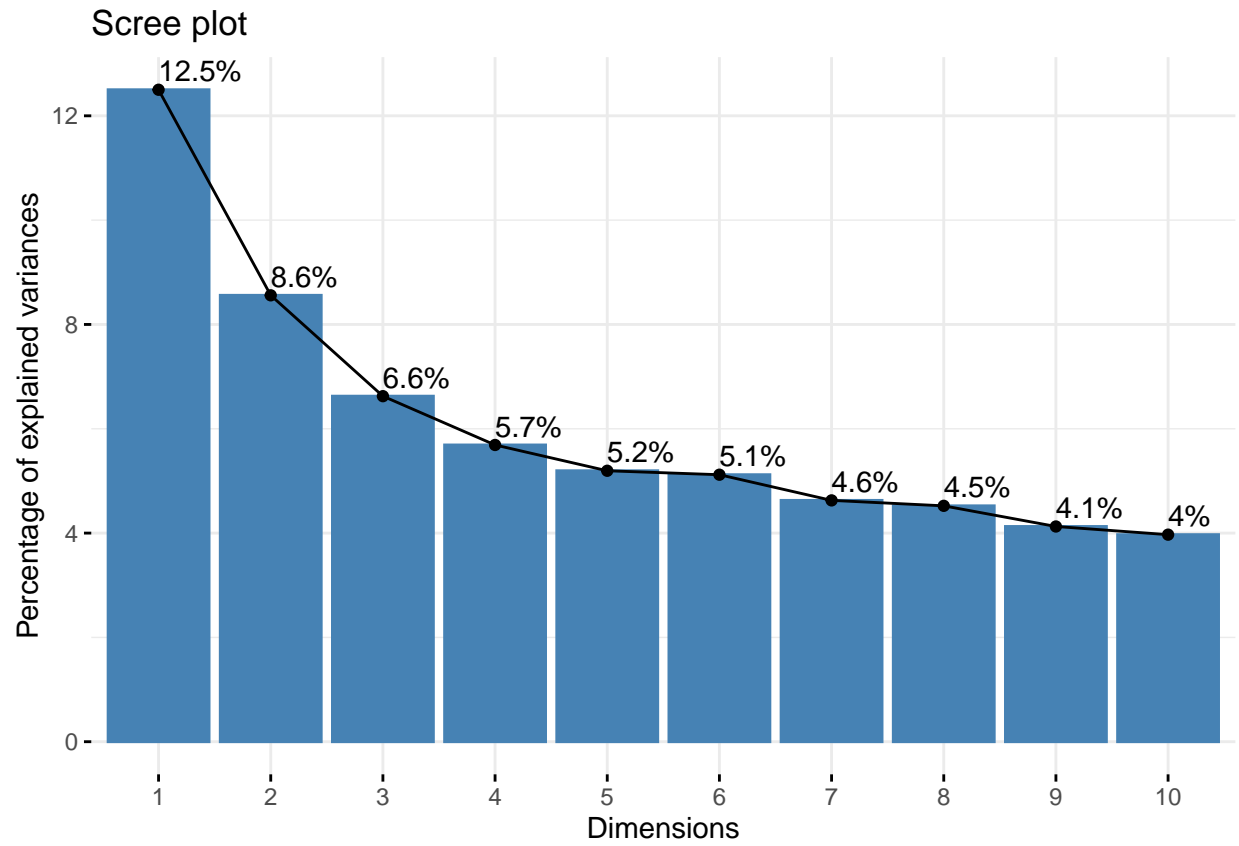
**Variable Categories** Dimension 1 contrasts users who do not save recipes, do not have a primary recipe website, do not modify recipes, etc. with users who use various platforms like Pinterest, blogs, videos, and online magazines to browse and discuss recipes. Basically users who do not often interact with online recipe content are negatively weighted and users who do frequently interact with recipes on various platforms are positively weighted.

Dimension 2 contrasts users who use Reddit to discuss and search recipes, often discuss recipes, and speak to their family and friends when browsing with users who use various platforms like Pinterest, blogs, videos, and online magazines to browse and discuss recipes, and who do not modify recipes. In other words, users who enjoy a more personal community/discussion based experience are negatively weighted, and users who prefer to use professionally developed recipes as written.

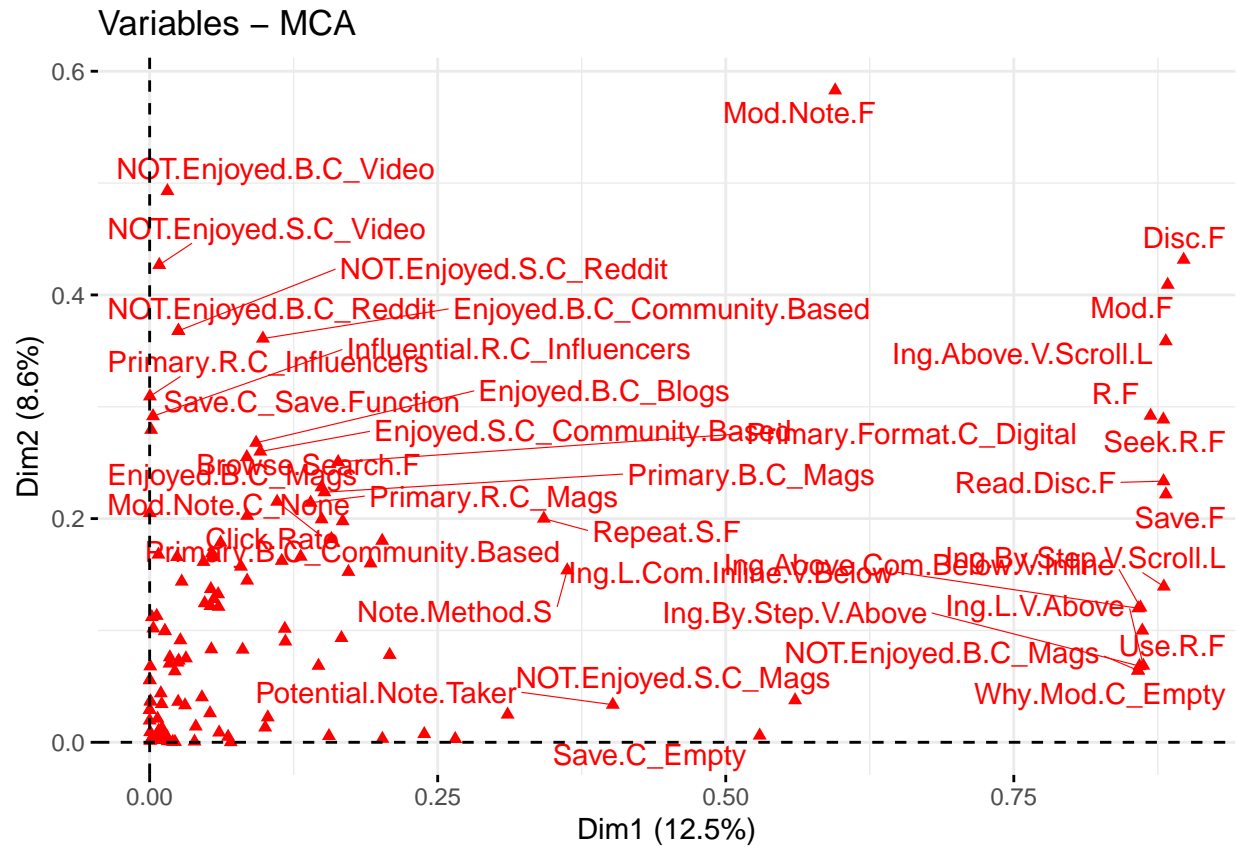
```
cleaned.Diet.No<-filter(search.data, Dietary.Restriction == "No")
cleaned.Diet.No<-cleaned.Diet.No%>%select(-c(Dietary.Restriction))
options(ggrepel.max.overlaps = 20)

cleaned.data<-mca.plot(cleaned.Diet.No)
```

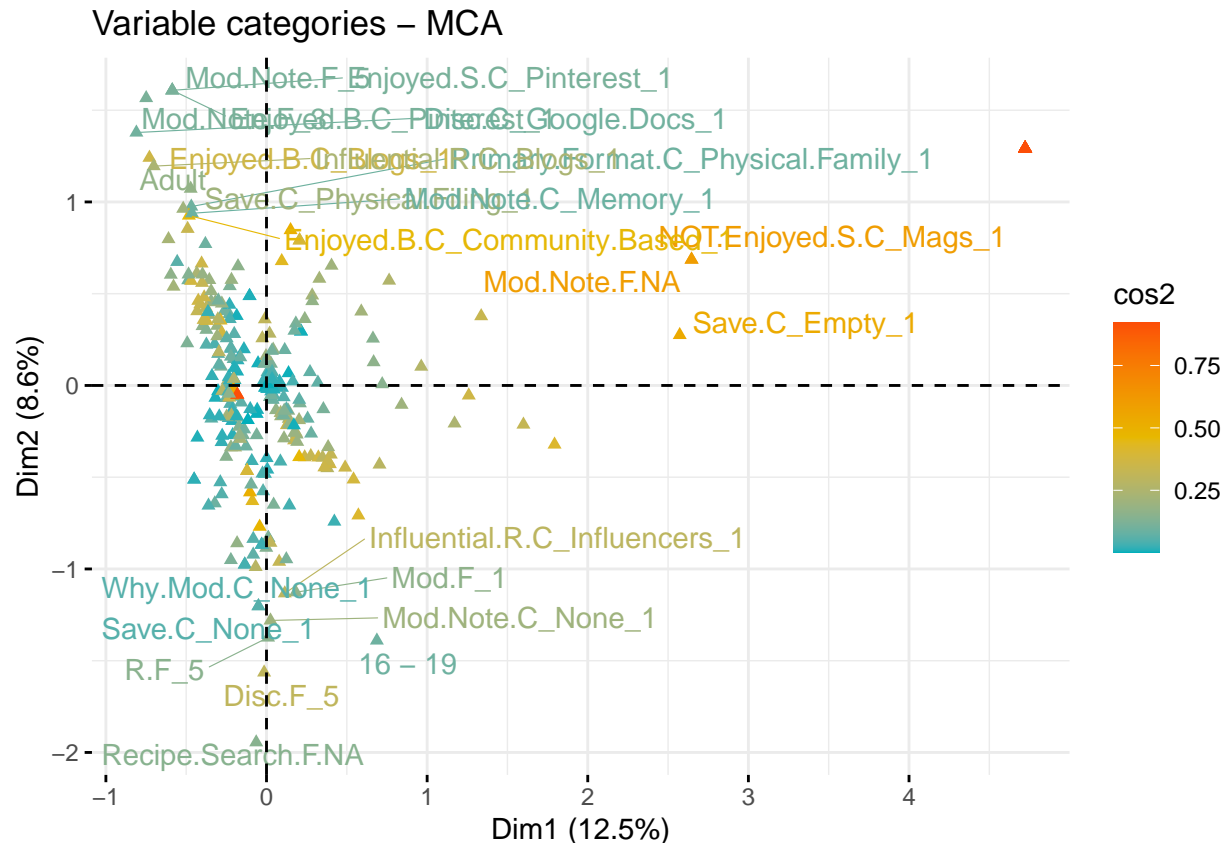




```
## Warning: ggrepel: 86 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```



```
## Warning: ggrepel: 279 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



#### #### Scree Plot

Using the elbow rule, I will be looking at the first two components of this MCA analysis. Therefore this plot will only capture 21.1% of the variability in the data. Slightly More variance is captured in the components when we stratify by diet.

**Variables** Recipe discussion frequency, modification frequency, ingredients listed above vs. ingredients pinned left while scrolling, review frequency, save frequency, and reading discussion frequency are all well represented in dimension 1.

Not enjoying video for searching or browsing is well represented in dimension 2.

**Variable Categories** The red point in the top left is all the questions answered with NA.

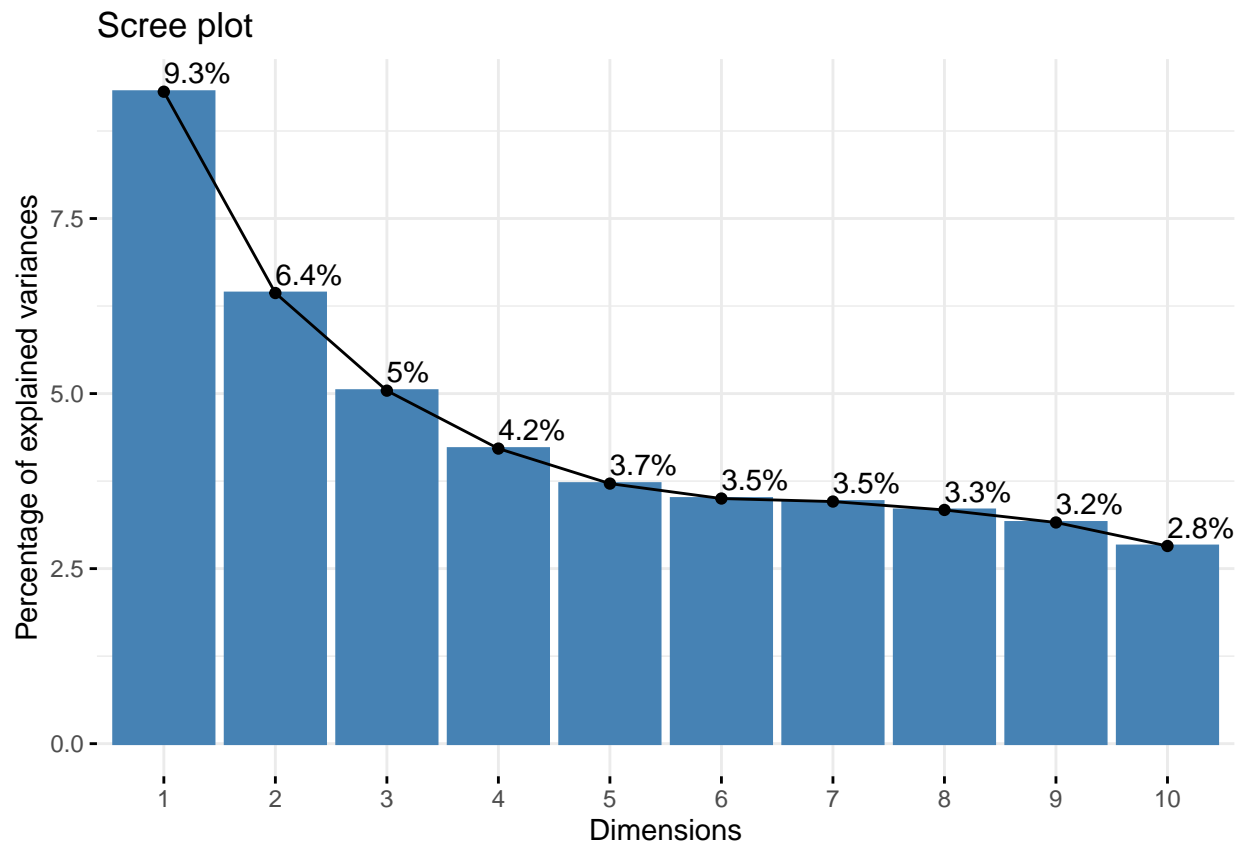
Dimension 1 contrasts users who left questions blank with users who answered survey questions.

Dimension 2 contrasts users who frequently review and discuss recipes, don't have a save method, don't have a modification motivation, do not frequently modify recipes, click on many recipes before cooking, and follow influencers recommendations with users who frequently take note of their modification, save recipes, and enjoy blogs, magazines and family/friends for browsing, searching, and receiving recommendations. In other words, adults users are positively weighted and associated with physical filing, frequent note taking, and using well recognized sources for recipes and reviews, while Users who do not modify recipes often, organize their recipe saves, consult influencers recommendations, and click through/browse many recipes are negatively weighted.

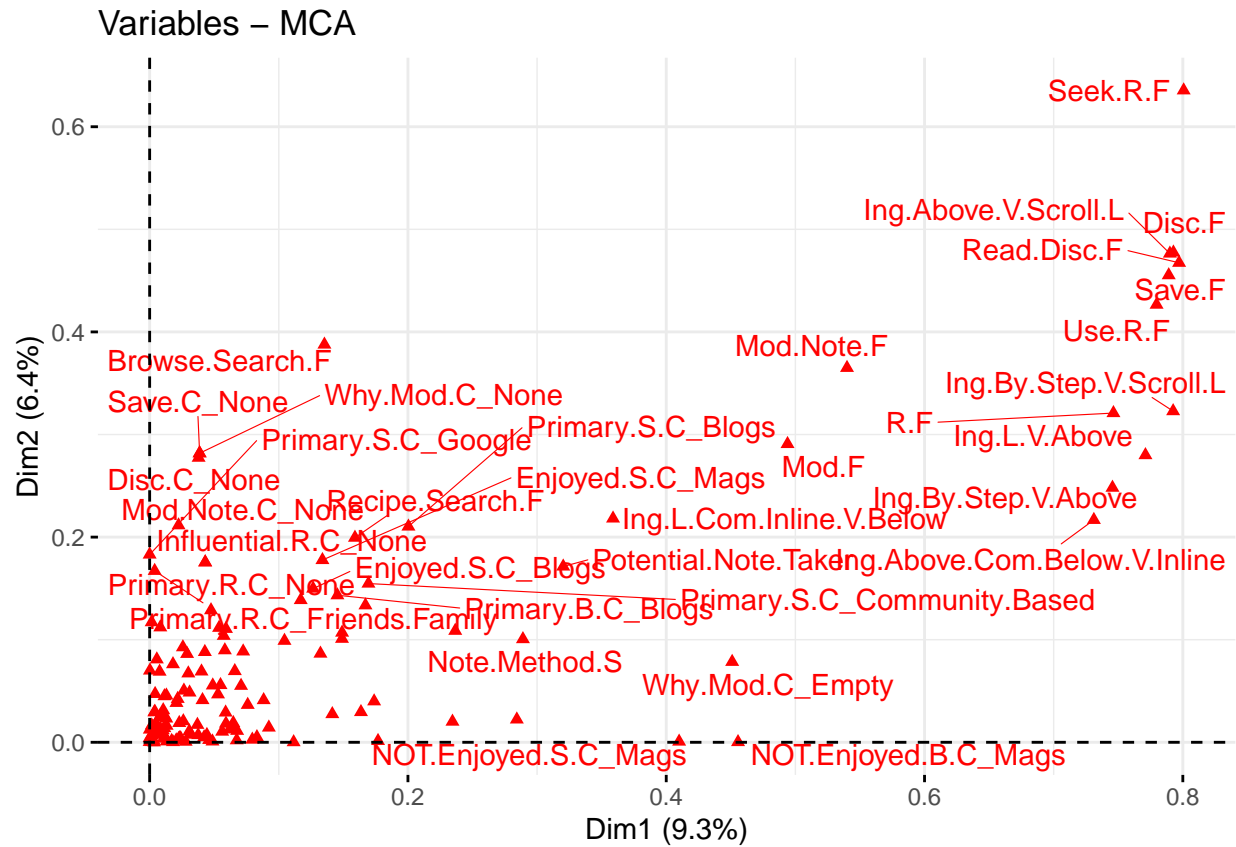
## Stratify by Age

```
cleaned.YA<-filter(search.data, Age == "YA")
cleaned.YA<-cleaned.YA%>%select(-c(Age))
options(ggrepel.max.overlaps = 22)

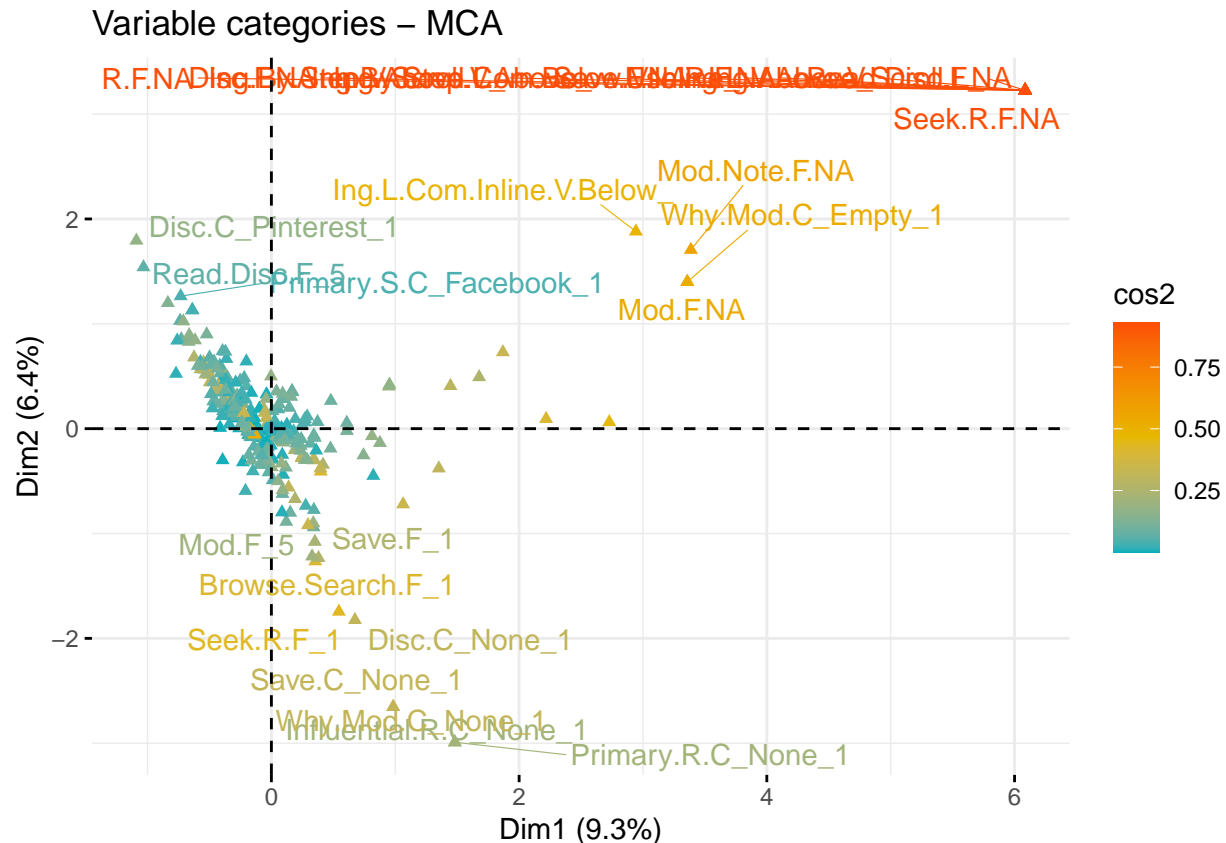
cleaned.data<-mca.plot(cleaned.YA)
```



```
## Warning: ggrepel: 105 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
## Warning: ggrepel: 307 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



#### #### Scree Plot

Using the elbow rule, I will be looking at the first two components of this MCA analysis. Therefore this plot will only capture 15.7% of the variability in the data. Slightly More variance is captured in the components when we stratify by age.

**Variables** Recipe discussion frequency, using review frequency, ingredients listed above vs. ingredients pinned left while scrolling, ingredients listed left vs above, review frequency, save frequency, and reading discussion frequency are all well represented in dimension 1.

Seeking review review frequency is well represented in both dimension.

**Variable Categories** The red point in the top left is all the questions answered with NA.

Dimension 1 contrasts users who left questions blank with users who answered survey questions.

Dimension 2 contrasts users who frequently discuss recipes, use Facebook to search, and use pinterest to browse with users who do not regularly interact with recipes online.

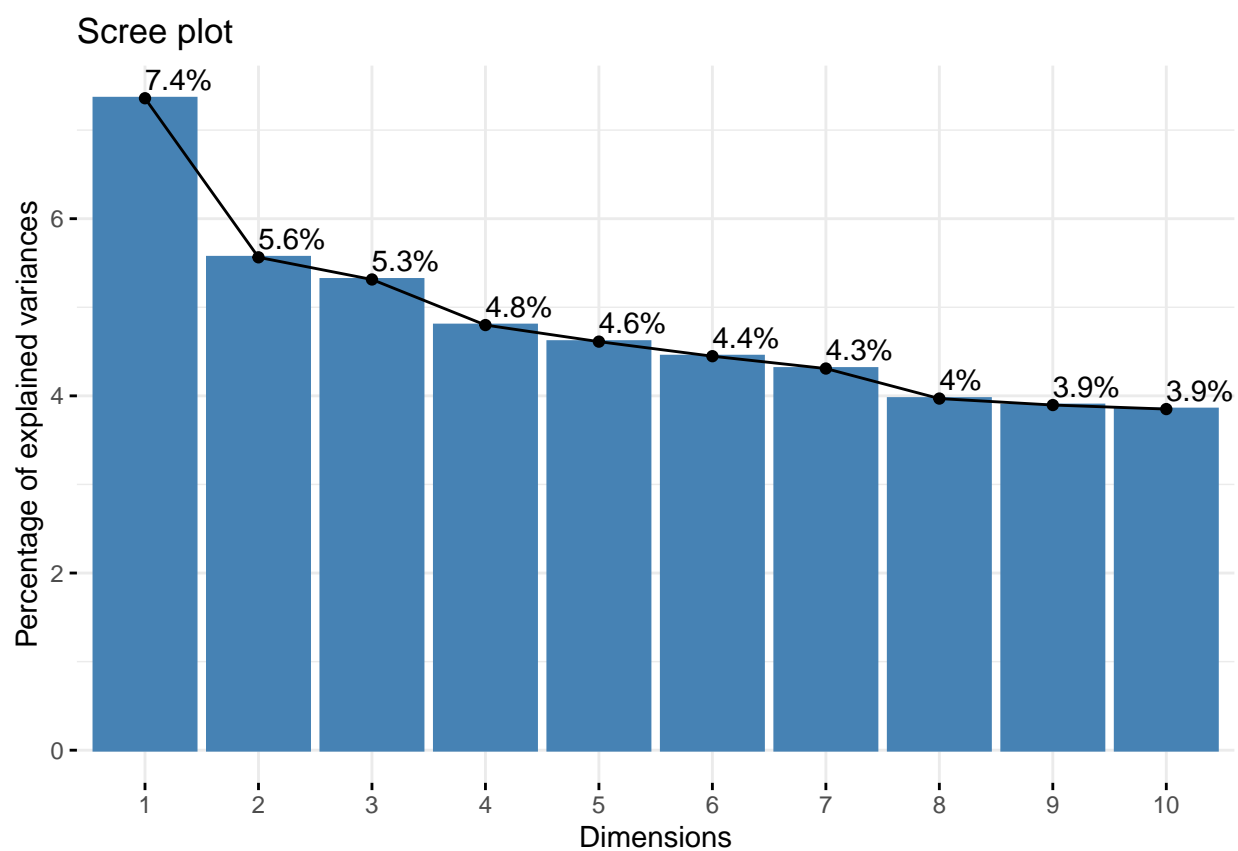
```
cleaned.adult<-filter(search.data, Age == "Adult")
cleaned.adult<-cleaned.adult%>%select(-c(Age))
nrow(cleaned.adult)
```

```
## [1] 6
```

This strata only has 6 sample units, so I will skip it in my further analysis. This strata has an atrificially low variance due to its low membership.

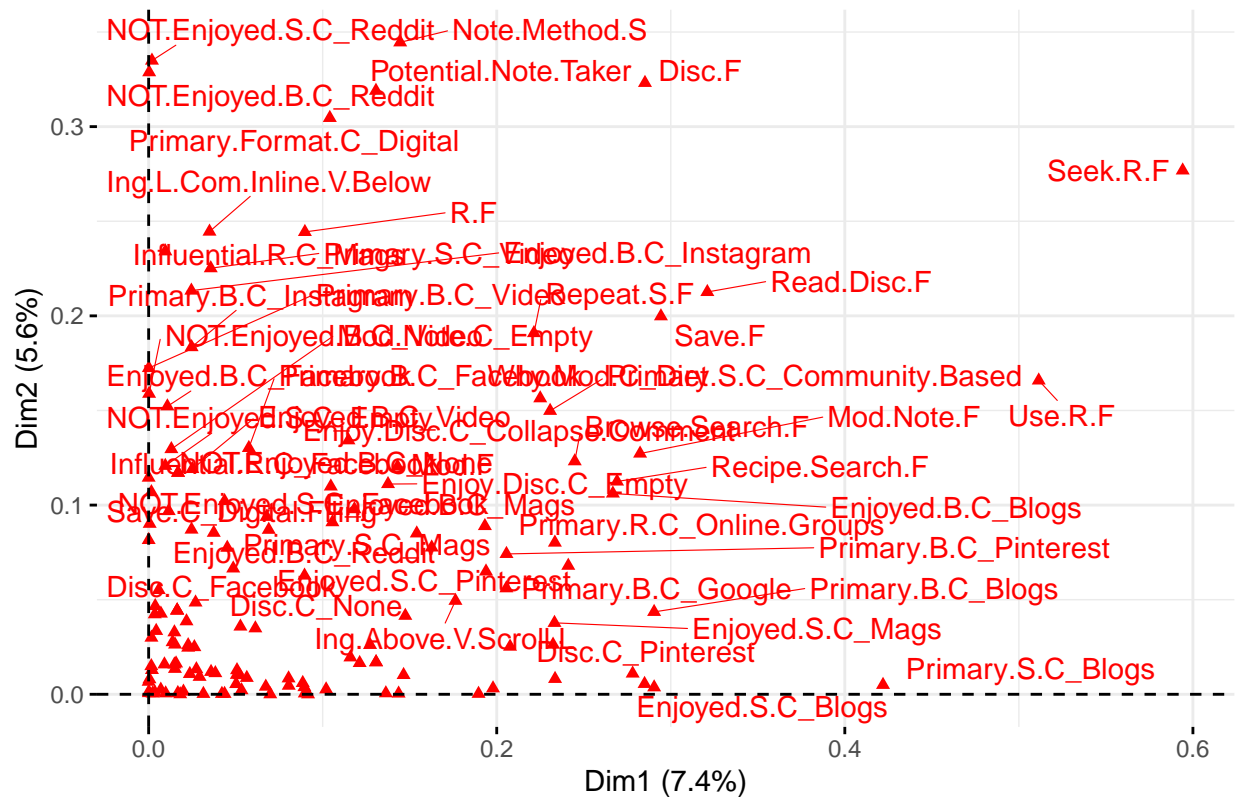
## Stratify by Who Cooks

```
primary.cook<-filter(search.data, Meal.Prepper == "Respondent")
primary.cook<-primary.cook%>%select(-c(Meal.Prepper))
cleaned.data<-mca.plot(primary.cook)
```



```
## Warning: ggrepel: 87 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

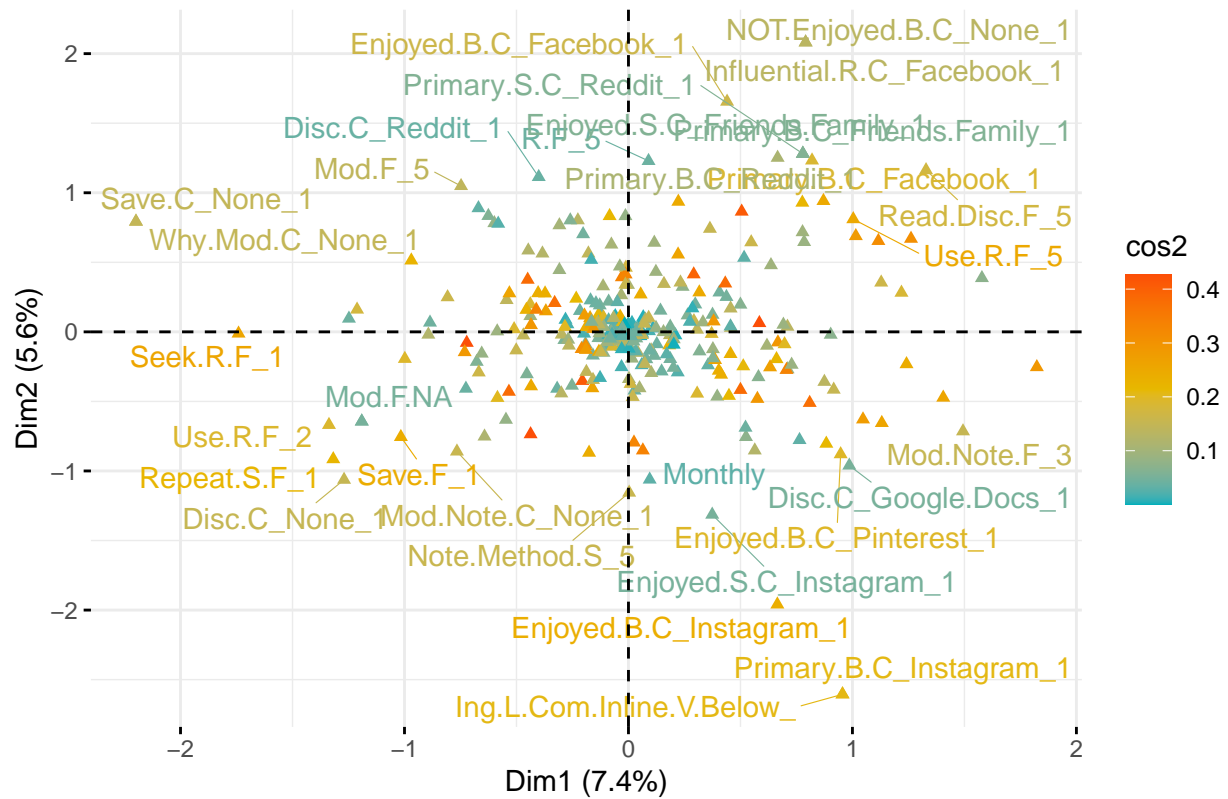
## Variables – MCA



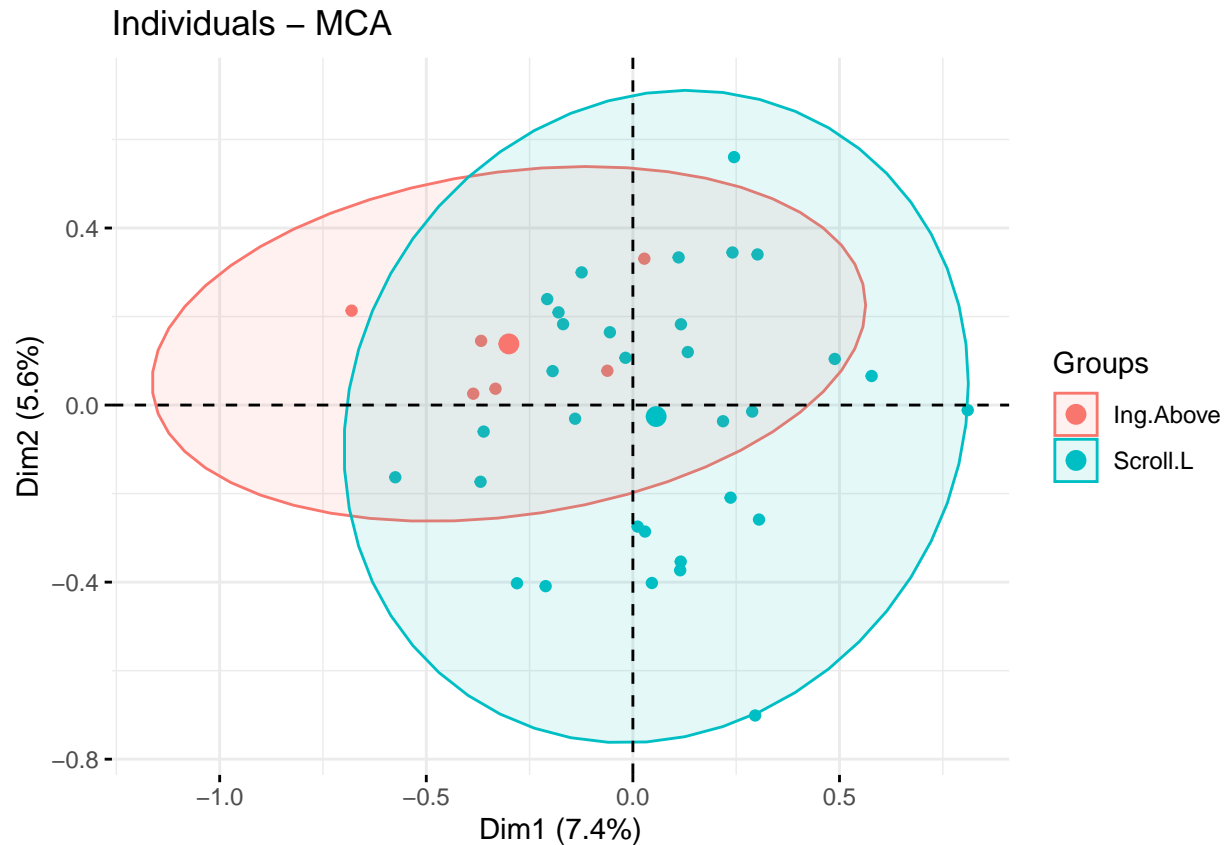
```
## Warning: ggrepel: 289 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



## Variable categories – MCA



```
fviz_mca_ind(cleaned.data, label="none", habillage = primary.cook$Ing.Above.V.Scroll.L, addEllipses = T
```



#### #### Scree Plot

Using the elbow rule, I will be looking at the first two components of this MCA analysis. Therefore this plot will only capture 13% of the variability in the data. Slightly less variance is captured in the components when we stratify by who cooks.

**Variables** Using review frequency are very well represented in dimension 1.

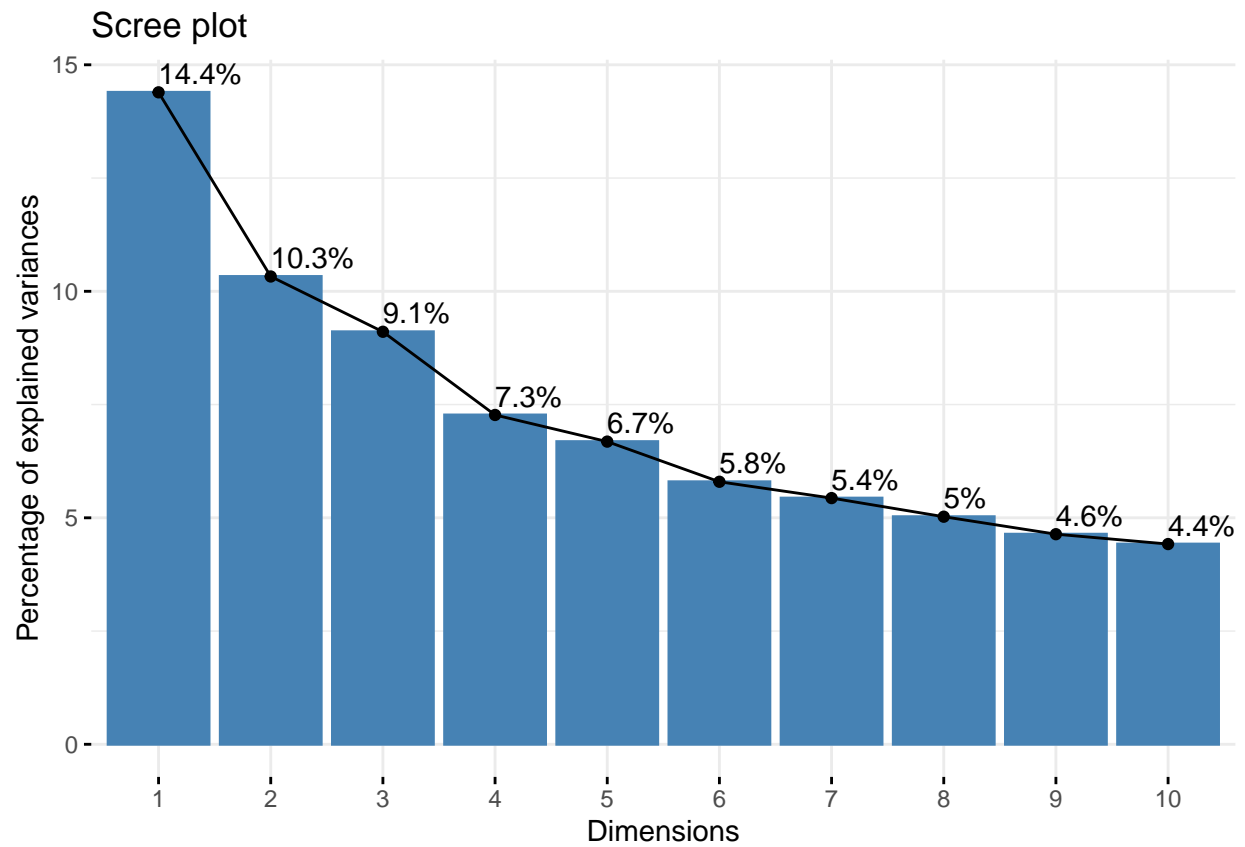
Not enjoying Reddit for searching or browsing, discussion frequency, and using digital recipes are well represented in dimension 2.

Seeking review review frequency is well represented in both dimension.

**Variable Categories** Dimension 1 contrasts users who do not frequently interact with recipes with users who frequently discuss recipes, take notes and use pinterest to discuss recipes.

Dimension 2 contrasts users that enjoy Instagram for browsing and searching, with users who enjoy Facebook and Reddit.

```
primary.cook.o<-filter(search.data, Meal.Prepper == "Other")
primary.cook.o<-primary.cook.o%>%select(-c(Meal.Prepper))
cleaned.data<-mca.plot(primary.cook.o)
```



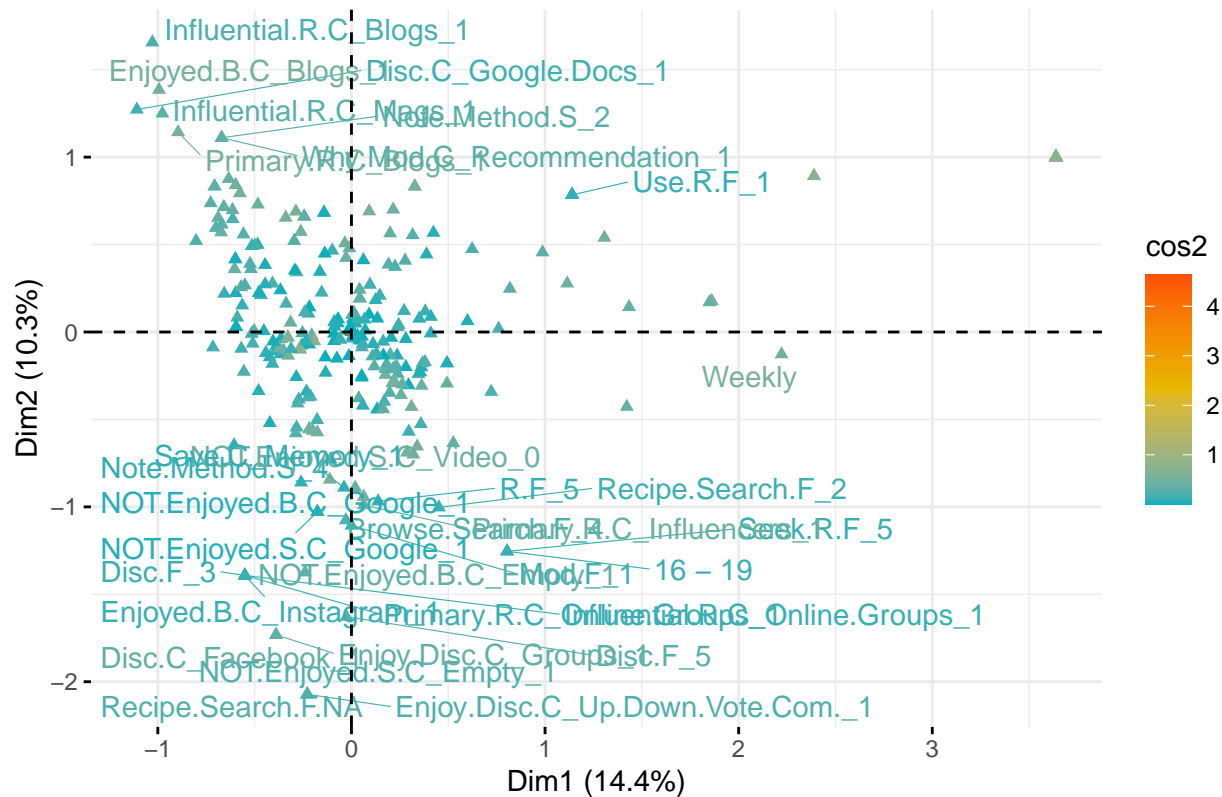
```
## Warning: ggrepel: 49 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

## Variables – MCA



```
## Warning: ggrepel. 258 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

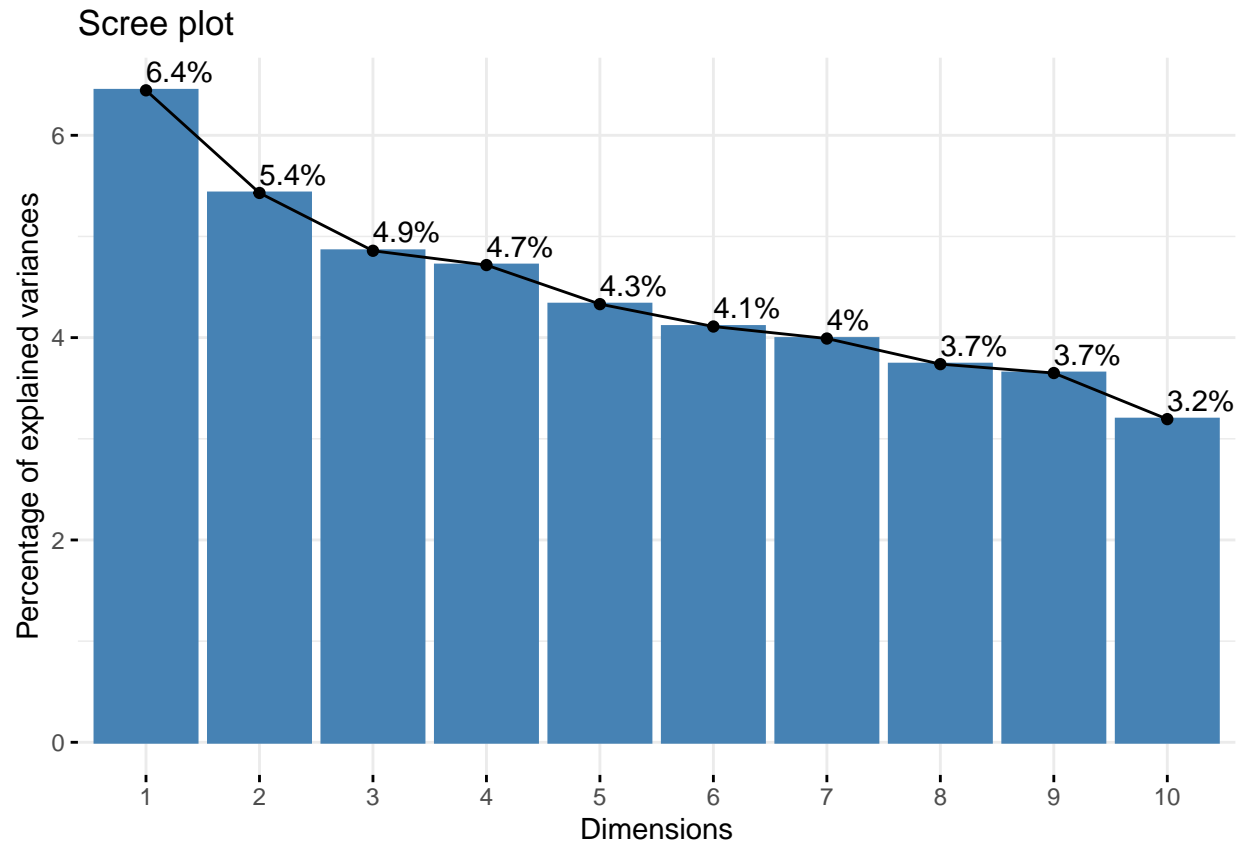
## Variable categories – MCA



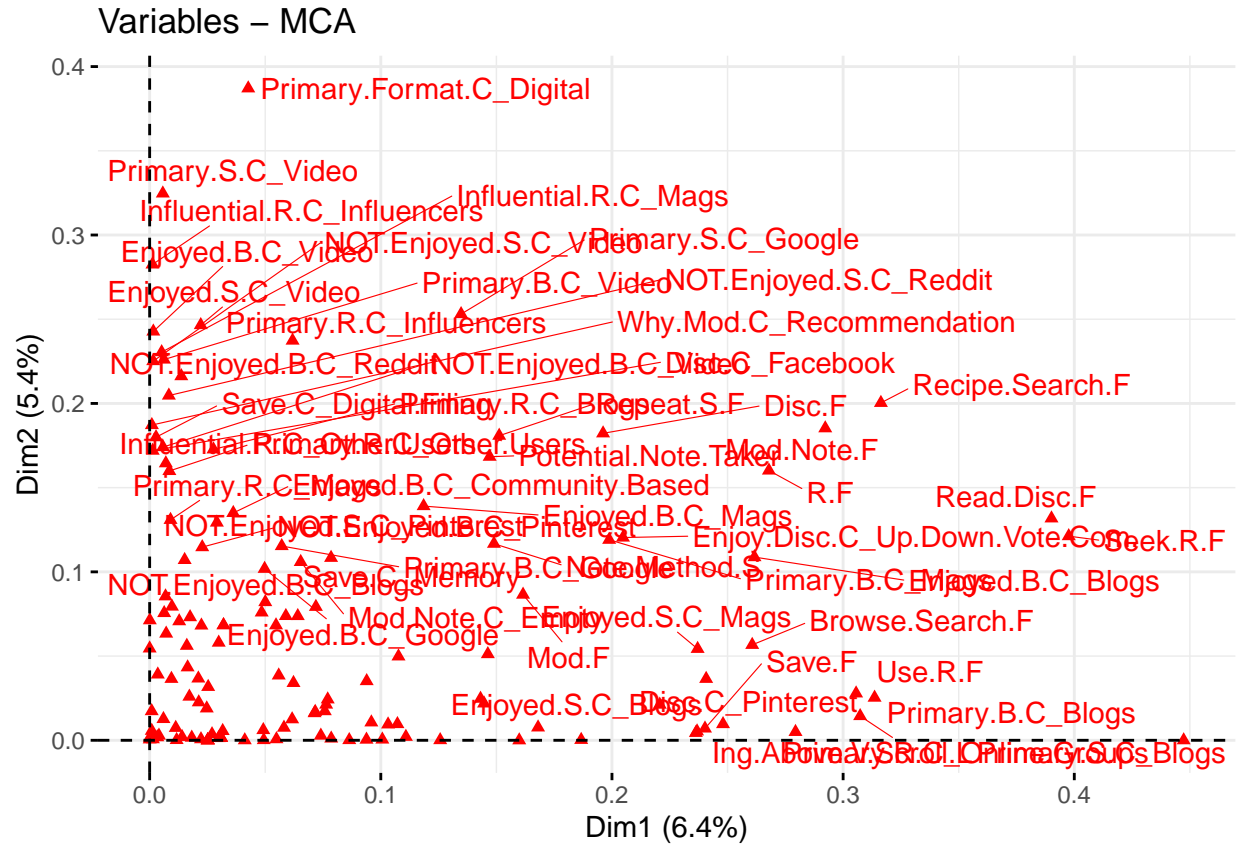
Looking at the Biplot, none of the variable categories have a high  $\cos^2$  value, so they are not well represented by the components. Therefore, further analysis will not be significant and should be interpreted with caution. ## Stratify By Cooking Frequency

## Overview

```
daily.cooks<-filter(search.data, Home.Cook.Rate == "Daily")
daily.cooks<-daily.cooks%>%select(-c(Home.Cook.Rate))
cleaned.data<-mca.plot(daily.cooks)
```

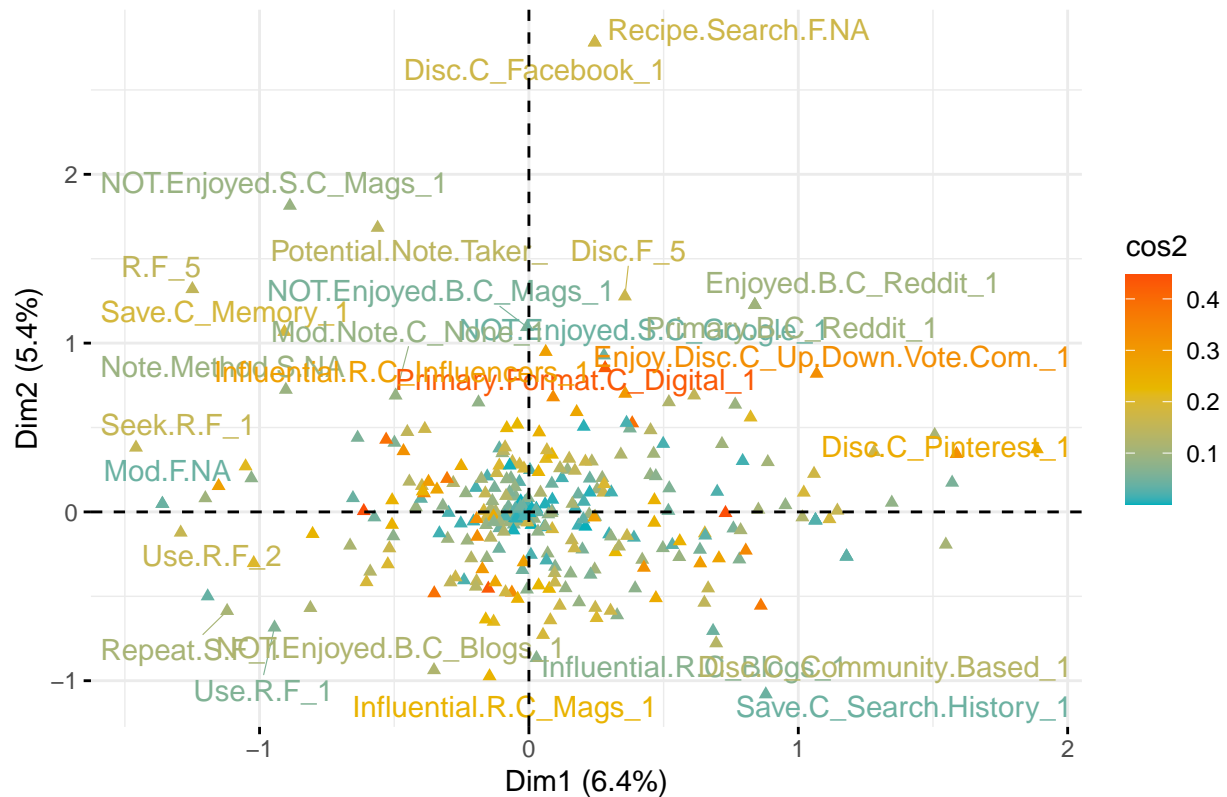


```
## Warning: ggrepel: 85 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
## Warning: ggrepel: 289 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

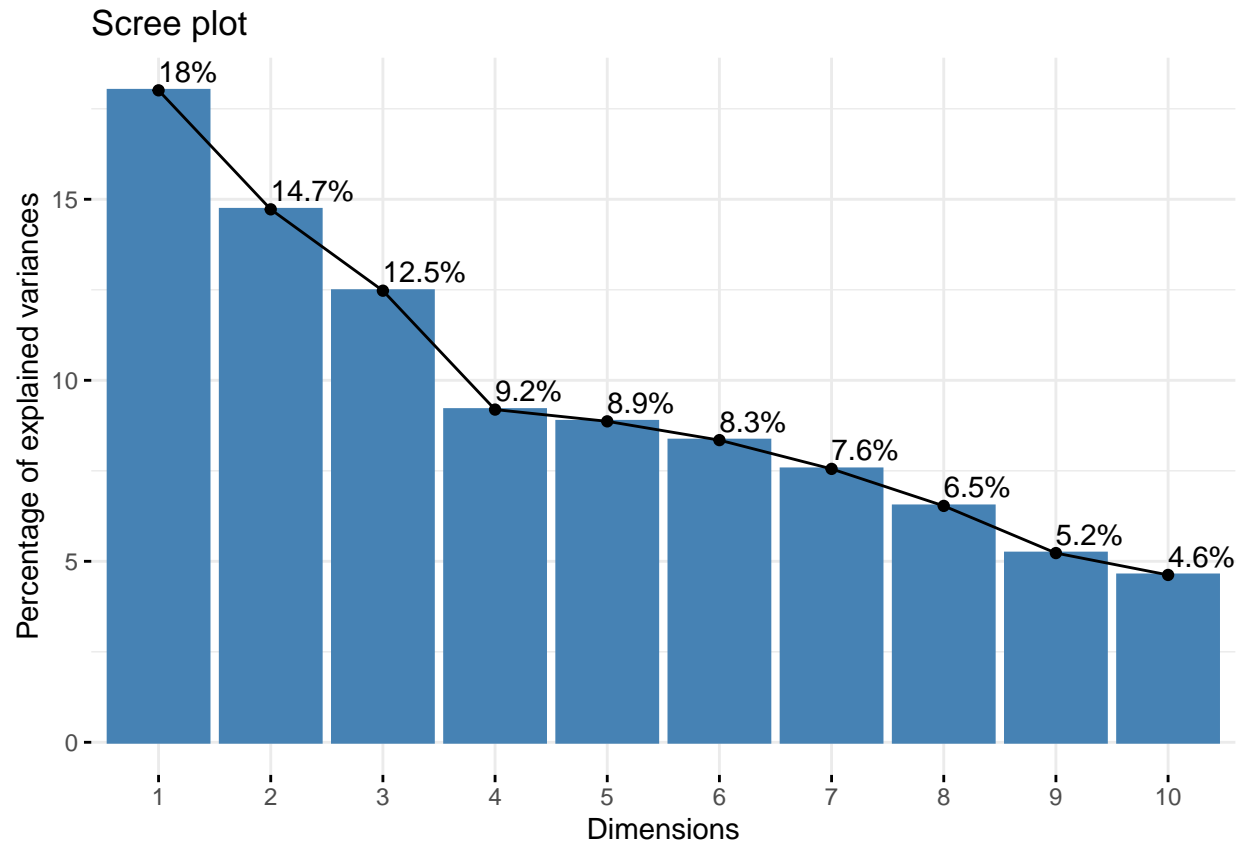
## Variable categories – MCA



Only 10% of the variance is explained by the components using the elbow rule, therefore this analysis is not as significant as the others I am considering.

```
weekly.cooks<-filter(search.data, !(Home.Cook.Rate == "Daily") )
weekly.cooks<-weekly.cooks%>%select(-c(Home.Cook.Rate))
mca.plot(weekly.cooks)
```



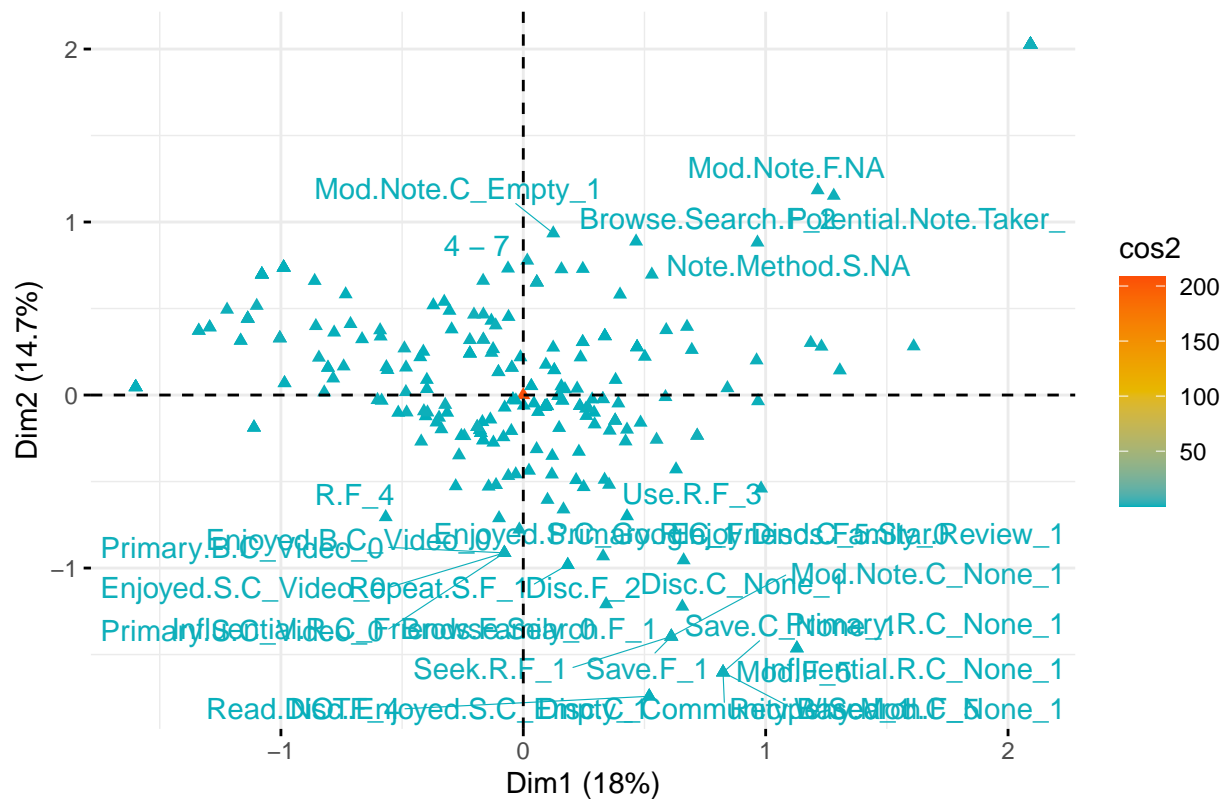


```
## Warning: ggrepel: 62 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

Figure 10: Biplot of the first two principal components (Dim1 and Dim2) for the 2008 dataset. The x-axis is Dim1 (18%) and the y-axis is Dim2 (12%). The plot shows the distribution of variables (red triangles) and their corresponding labels. A dashed vertical line is at Dim1 = 0.0. Variables like 'Save.F' and 'Seek.R.F' are in the top right, while 'Enjoyed.S.C\_Google' and 'Enjoyed.B.C\_Video' are in the top left. Many variables are clustered near the bottom of the plot.

34

## Variable categories – MCA



```
## **Results of the Multiple Correspondence Analysis (MCA)**
## The analysis was performed on 12 individuals, described by 110 variables
## *The results are available in the following objects:
##
##   name          description
## 1  "$eig"        "eigenvalues"
## 2  "$var"        "results for the variables"
## 3  "$var$coord"  "coord. of the categories"
## 4  "$var$cos2"   "cos2 for the categories"
## 5  "$var$contrib" "contributions of the categories"
## 6  "$var$v.test" "v-test for the categories"
## 7  "$ind"        "results for the individuals"
## 8  "$ind$coord"  "coord. for the individuals"
## 9  "$ind$cos2"   "cos2 for the individuals"
## 10 "$ind$contrib" "contributions of the individuals"
## 11 "$call"       "intermediate results"
## 12 "$call$marge.col" "weights of columns"
## 13 "$call$marge.li"  "weights of rows"
```

```
nrow(weekly.cooks)
```

```
## [1] 12
```

Looking at the Biplot, none of the variable categories have a high cos2 value, so they are not well represented by the components. Therefore, further analysis will not be significant and should be interpreted with caution.