TP1 - Data Analysis I

Guerzoniansus

2024-04-15

Contents

Résumé

L'objectif de ce projet est d'appliquer les outils que nous avons étudiés pendant le cours du logiciel statistique R, dans le cas d'une étude de cas réelle. Les codes utiliser pour répondre aux questions sont intégrés dans le corps du rapport grâce à R Markdown.

Importation des packages nécessaires et préparation de l'environnement R

```
library(dplyr)
library(readxl)
library(here)
library(labelled)
```

```
rm(list=ls())
```

```
setwd(dirname(getwd())) # récupère le dossier dans lequel se trouve le script R, prend le dossier paren
c_dir <- getwd()
```

Recodage et labelisation

Description de la base

Le fichier Base_Projet.xlsx contient 250 observations et 33 variables. La première colonne key correspond à l'identifiant de la PME. Les variables sont les suivantes :

sexe: Sexe

age: Age

sit_mat: Situation maritale

si_chef_men: Statut dans le ménage

ethnie: Ethnie

occupation: Occupation

formation: Formation

niveau_alphabs: Niveau d'alphabétisation

sup_tot: Quelle est la superficie emblavée en sésame la saison passée ? en hectares

types_varietes: Quelles sont les variétés que vous utilisez pour la production de sésame ?

var_trad: Si variétés traditionnelles, donnez les noms

raison_var_trad: Pourquoi utilisez vous les variétés traditionnelles ?

var_amel: Si variétés améliorées, laquelle utilisez-vous ?

raison_var_amel: Pourquoi utilisez vous les variétés améliorées ?

criteres_var: Quelles sont les critères de choix des variétés de sésame ?

Importation et mise en forme

Importation de la base de données dans un objet de type data.frame nommé base_tp2

```
base_tp2 <- read_excel(paste0(c_dir, "/data/Base TP2.xlsx"))
```

Nombre de lignes (i.e. le nombre de PME) et de colonnes (i.e. nombre de variables) de la base projet

```
n_rows <- nrow(base_tp2)
n_cols <- ncol(base_tp2)
cat("Nombre de PME (lignes): ", n_rows, "\n")
```

```
## Nombre de PME (lignes):  53
```

```
cat("Nombre de variables (colonnes): ", n_cols, "\n")
```

```
## Nombre de variables (colonnes):  30
```

Recodage et labelisation

```r
base_tp2$sexe <- labelled::labelled(
  base_tp2$sexe,
  c(`Homme` = 1,
    `Femme` = 2))

base_tp2$sit_mat <- labelled::labelled(
  base_tp2$sit_mat,
  c(`Marié(e)` = 1,
    `Veuf(ve)` = 3,
    `Divorcé(e)` = 4,
    `Séparé(e)` = 5,
    `Célibataire` = 6))

base_tp2$si_chef_men <- labelled::labelled(
  base_tp2$si_chef_men,
  c(`Femme du chef de ménage` = 1,
    `Chef de ménage` = 2,
    `Fils-fille du chef de ménage` = 3,
    `Autres` = 99))

base_tp2$ethnie <- labelled::labelled(
  base_tp2$ethnie,
  c(`Wolof` = 1,
    `Pulaar/Toucouleur` = 2,
    `Sérère` = 3,
    `Mandika/Bambara` = 4,
    `Soninké` = 5,
    `Diola` = 6,
    `Manjack` = 7,
    `Bainouk` = 8,
    `Maures` = 9,
    `Balante` = 10,
    `Autre` = 77))

base_tp2$occupation <- labelled::labelled(
  base_tp2$occupation,
  c(`Agriculture, Elevage, Sylviculture, Pêche` = 1,
    `Activités extractives` = 2,
    `Activités de fabrication (Artisanat)` = 3,
    `Activités de transformation` = 4,
    `Production et distribution d'électricité et de gaz` = 5,
    `Production et distribution d'eau, assainissement, traitement des déchets et dépollution` = 6))

base_tp2$formation <- labelled::labelled(
  base_tp2$formation,
  c(`Non scolarisé` = 1,
    `Elémentaire` = 2,
    `Moyen` = 3,
    `Secondaire` = 4,
    `Licence` = 5,
    `Master` = 6,
    `Doctorat` = 7,
    `Ne sait pas` = 99))
```

```
base_tp2$niveau_alphabs <- labelled::labelled(
  base_tp2$niveau_alphabs,
  c(`Sans niveau` = 0,
    `Sait lire dans une langue` = 1,
    `Sait lire et écrire dans une langue` = 2))

base_tp2$types_varietes <- labelled::labelled(
  base_tp2$types_varietes,
  c(`Traditionnelles` = "1",
    `Améliorées` = "2"))

base_tp2$criteres_var <- labelled::labelled(
  base_tp2$criteres_var,
  c(`Rendements élevés` = "1",
    `Taille des graines` = "2",
    `Résistantes aux maladies/ravageurs` = "3",
    `Tolérantes aux sécheresses` = "4",
    `Tolérantes aux inondations` = "5",
    `Faible charge de travail` = "6",
    `Faibles quantités d'intrants` = "7",
    `Facile à transformer` = "8",
    `Haute teneur en huile` = "9",
    `Haut rendement après transformation` = "10",
    `Demande sur le marché` = "11",
    `Bon gôut` = "12",
    `Belle couleur` = "13",
    `Haut rendement en fourrages` = "14",
    `Qualité du fourrage` = "15",
    `Autres à spécifier` = "16"))
```

```
#check labels
expss::val_lab(base_tp2$sexe)
```

```
## Homme Femme
##     1     2
```

```
expss::val_lab(base_tp2$sit_mat)
```

```
##    Marié(e)    Veuf(ve)  Divorcé(e)   Séparé(e) Célibataire
##           1           3           4           5           6
```

```
expss::val_lab(base_tp2$si_chef_men)
```

```
##      Femme du chef de ménage               Chef de ménage
##                            1                            2
## Fils-fille du chef de ménage                       Autres
##                            3                           99
```

```
expss::val_lab(base_tp2$ethnie)
```

```
##             Wolof Pulaar/Toucouleur             Sérère   Mandika/Bambara
##                1                   2                  3                 4
##           Soninké             Diola           Manjack           Bainouk
##                5                   6                  7                 8
##            Maures           Balante             Autre
##                9                  10                 77
```

```
expss::val_lab(base_tp2$occupation)
```

```
##                                                Agriculture, Elevage, Sylviculture, Pêche
##                                                                                        1
##                                                                    Activités extractives
##                                                                                        2
##                                                Activités de fabrication (Artisanat)
##                                                                                        3
##                                                                 Activités de transformation
##                                                                                        4
##                                      Production et distribution d'électricité et de gaz
##                                                                                        5
## Production et distribution d'eau, assainissement, traitement des déchets et dépollution
##                                                                                        6
```

```
expss::val_lab(base_tp2$formation)
```

```
## Non scolarisé   Elémentaire         Moyen    Secondaire       Licence
##             1             2             3             4             5
##        Master      Doctorat   Ne sait pas
##             6             7            99
```

```
expss::val_lab(base_tp2$niveau_alphabs)
```

```
##                         Sans niveau           Sait lire dans une langue
##                                   0                                   1
## Sait lire et écrire dans une langue
##                                   2
```

```
expss::val_lab(base_tp2$types_varietes)
```

```
## Traditionnelles      Améliorées
##             "1"             "2"
```

```
expss::val_lab(base_tp2$criteres_var)
```

```
##                   Rendements élevés                  Taille des graines
##                                 "1"                                 "2"
##  Résistantes aux maladies/ravageurs          Tolérantes aux sécheresses
##                                 "3"                                 "4"
##          Tolérantes aux inondations            Faible charge de travail
##                                 "5"                                 "6"
##        Faibles quantités d'intrants                Facile à transformer
```

```
##                                "7"                                "8"
##               Haute teneur en huile Haut rendement après transformation
##                                "9"                               "10"
##               Demande sur le marché                         Bon gôut
##                               "11"                               "12"
##                       Belle couleur        Haut rendement en fourrages
##                               "13"                               "14"
##                 Qualité du fourrage                 Autres à spécifier
##                               "15"                               "16"
```

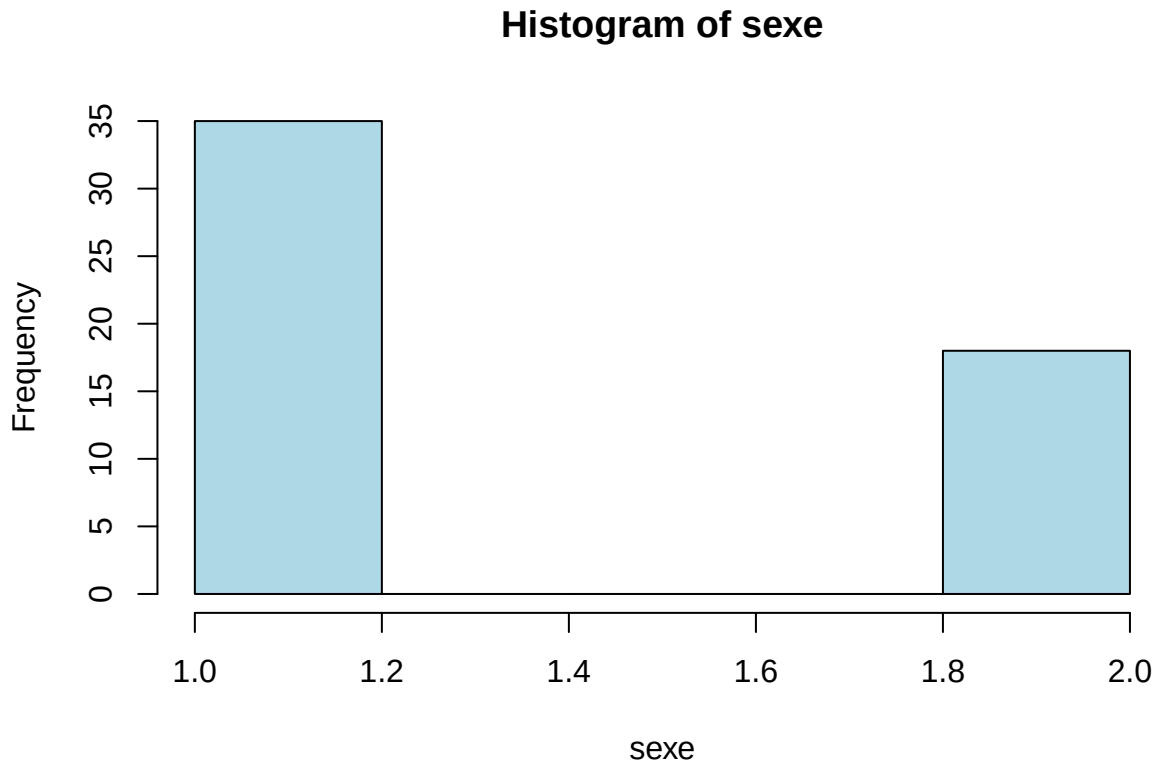
```
attach(base_tp2)
labelled::var_label(sexe) = "Sexe"

labelled::var_label(age) = "Age"

labelled::var_label(sit_mat) = "Situation maritale"

labelled::var_label(si_chef_men) = "Statut dans le ménage"

labelled::var_label(ethnie) = "Ethnie"

labelled::var_label(occupation) = "Occupation"

labelled::var_label(formation) = "Formation"

labelled::var_label(niveau_alphabs) = "Niveau d'alphabétisation"

# labelled::var_label(sup_tot) = "Quelle est la superficie totale emblavée en sésame la saison passée e


labelled::var_label(types_varietes) = "Quelles sont les variétés que vous utilisez pour la production d

#labelled::var_label(var_trad) = "Si variétés traditionnelles, donnez les noms ?"
#Cette variable n'existe pas dans la base

# labelled::var_label(raison_var_trad) = "Pourquoi utilisez-vous les variétés traditionnelles ?"
#Cette variable n'existe pas dans la base


# labelled::var_label(var_amel) = "Si variétés améliorées, laquelle utilisez-vous ?"
#Cette variable n'existe pas dans la base


# labelled::var_label(raison_var_amel) = "Pourquoi utilisez-vous les variétés améliorées ?"
#Cette variable n'existe pas dans la base


labelled::var_label(criteres_var) = "Quelles sont les critères de choix des variétés de sésame ?"
```

Analyse descriptive

```
source("script_p.R")
```

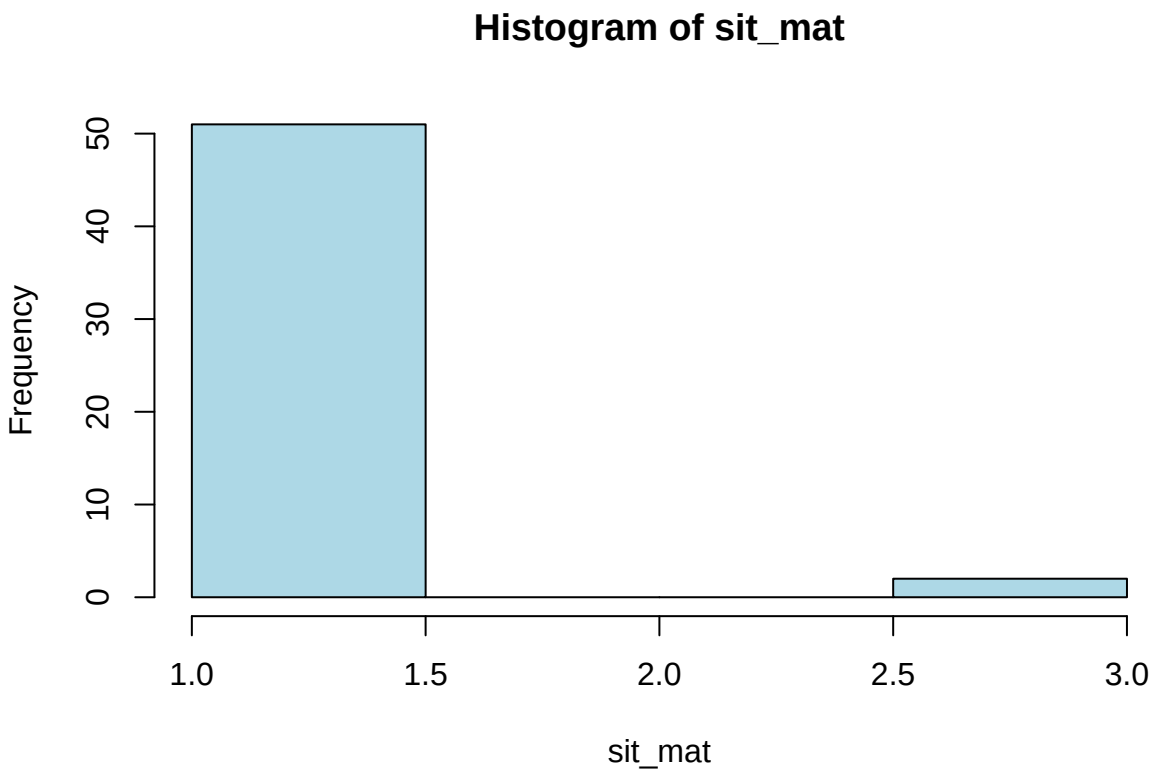
```
univarie(base_tp2, "sexe", plot = TRUE)
```

```
##
##  1  2
## 35 18
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    1.00    1.00    1.34    2.00    2.00
```



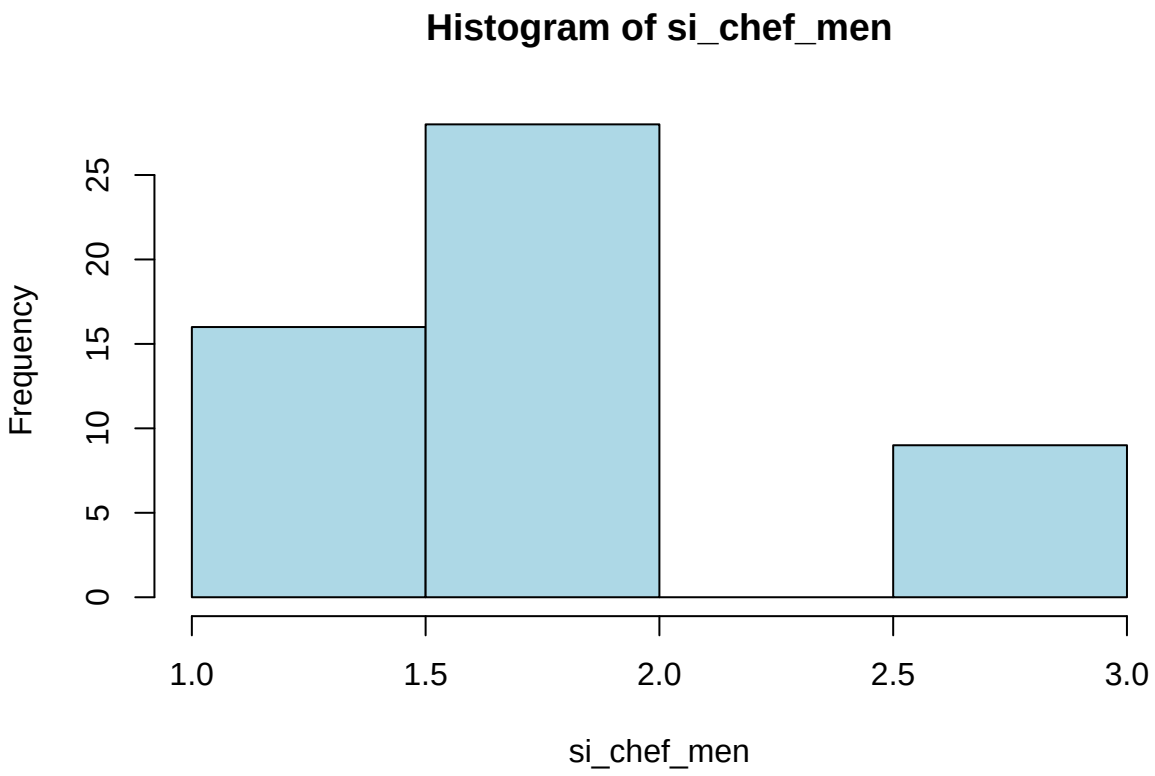

```
univarie(base_tp2, "sit_mat", plot = TRUE)
```

```
##
##  1  3
## 51  2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.075   1.000   3.000
```



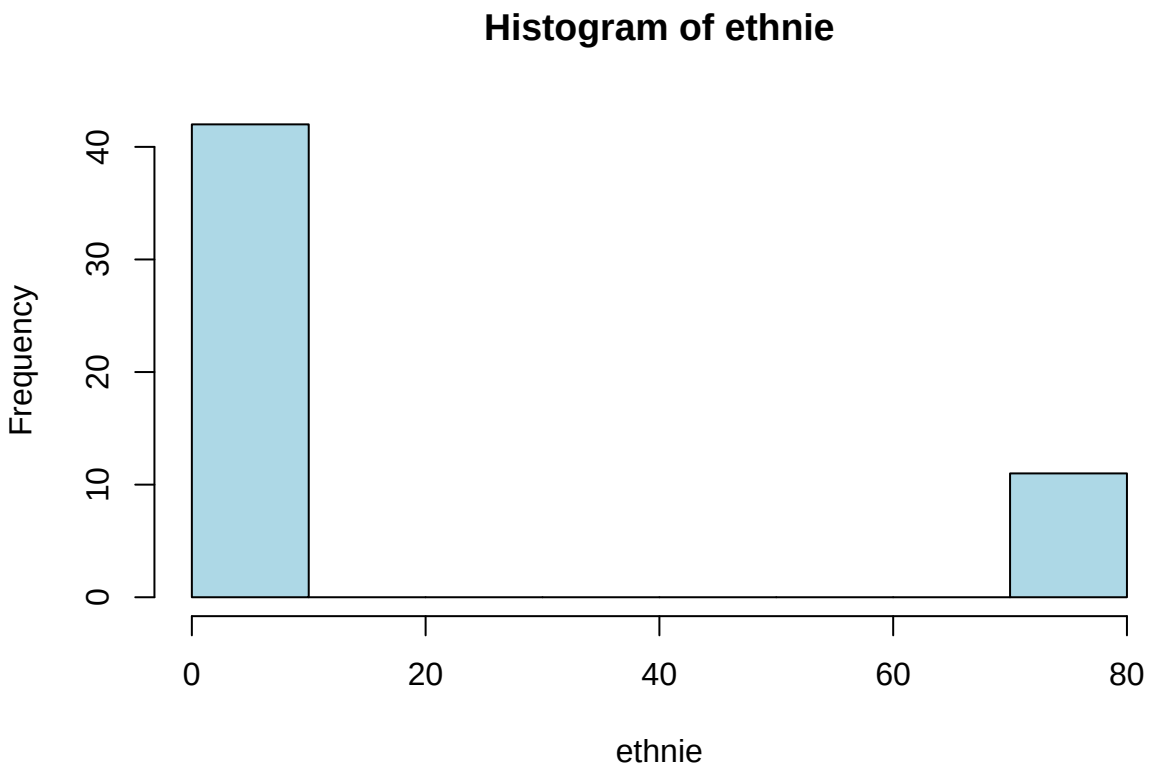

```
univarie(base_tp2, "si_chef_men", plot = TRUE)
```

```
##
##  1  2  3
## 16 28  9
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   1.868   2.000   3.000
```



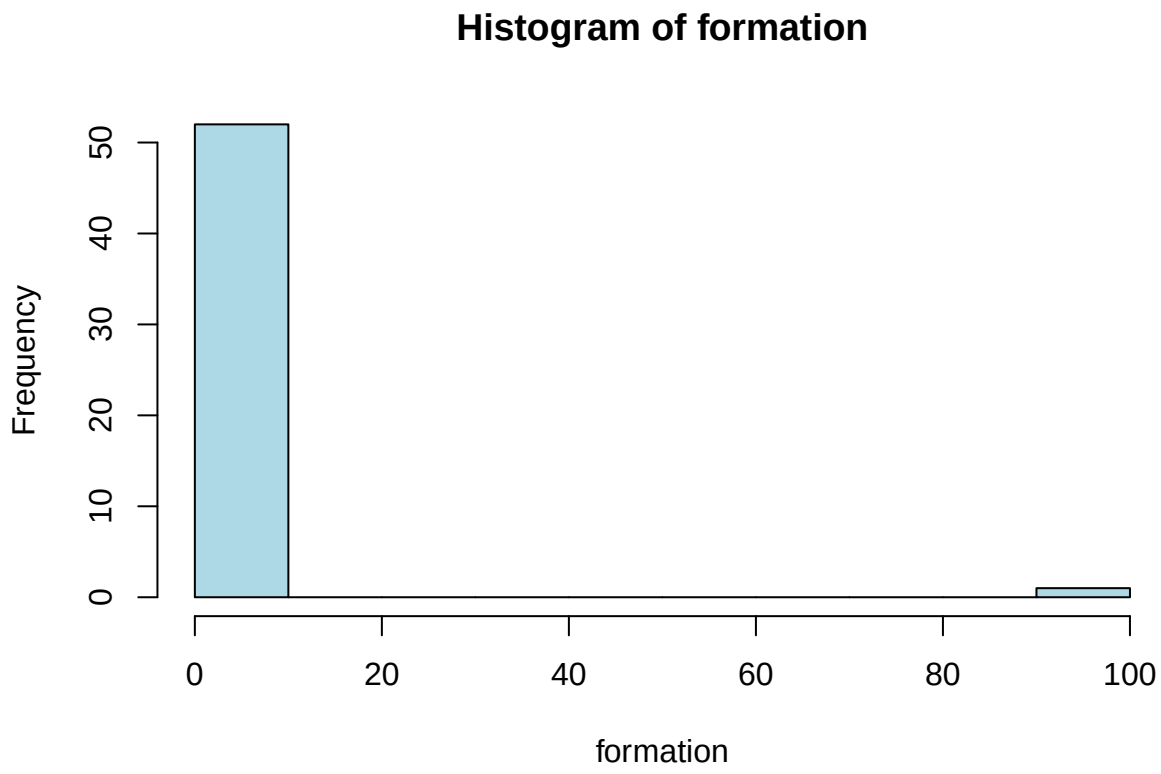

```
univarie(base_tp2, "ethnie", plot = TRUE)
```

```
##
##  1  2  3  4  6 10 77
## 10  7 13  8  1  3 11
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    2.00    3.00   18.45   10.00   77.00
```



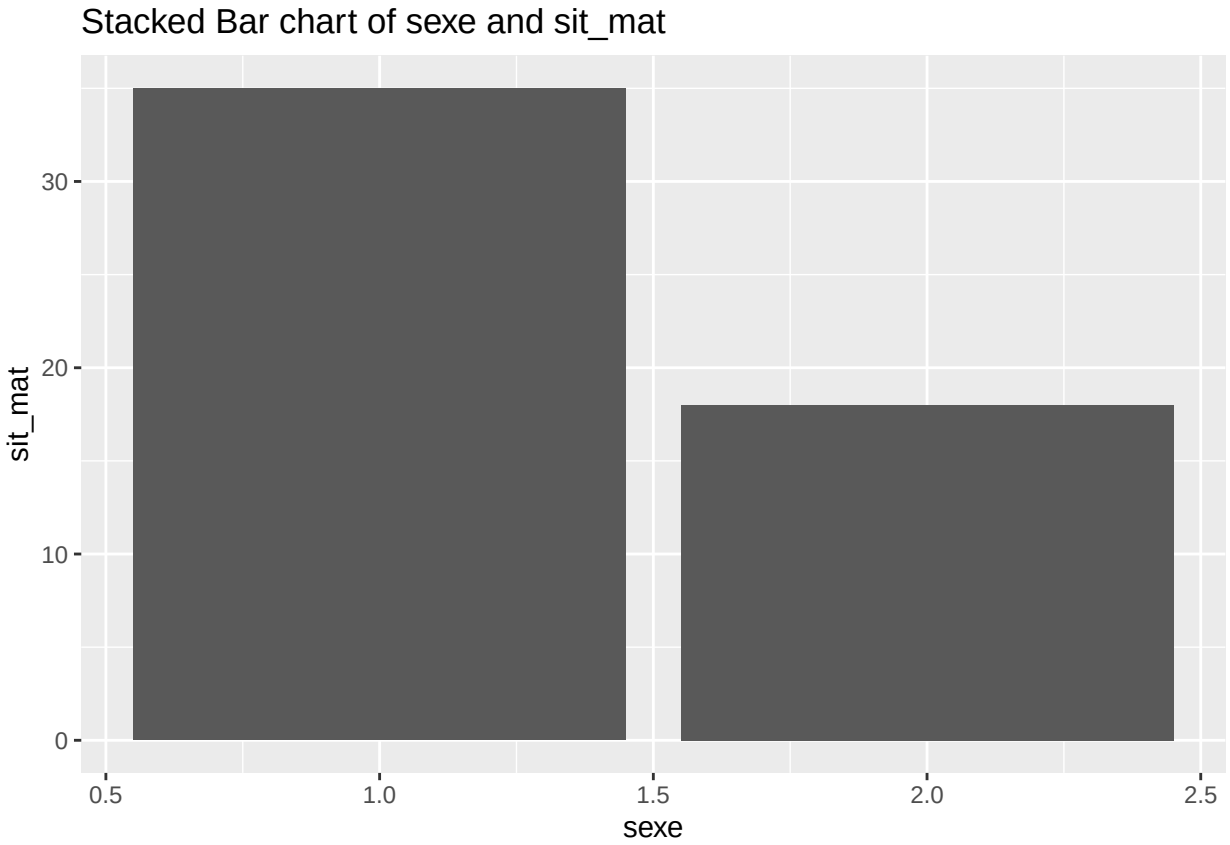

```
univarie(base_tp2, "formation", plot = TRUE)
```

```
##
##  1  2  3  4  5 99
## 30 10  7  3  2  1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   3.623   2.000  99.000
```

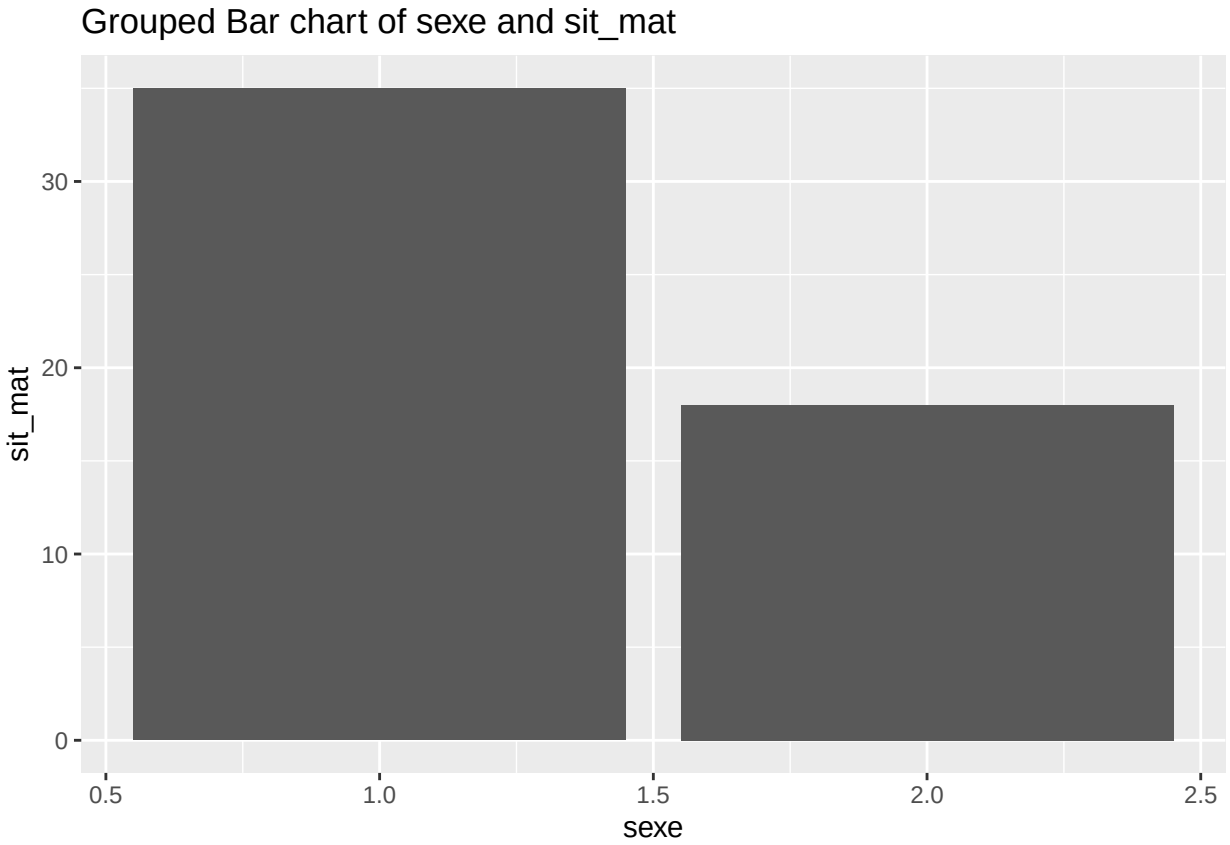



```
bivarie(base_tp2, "sexe", "sit_mat")
```

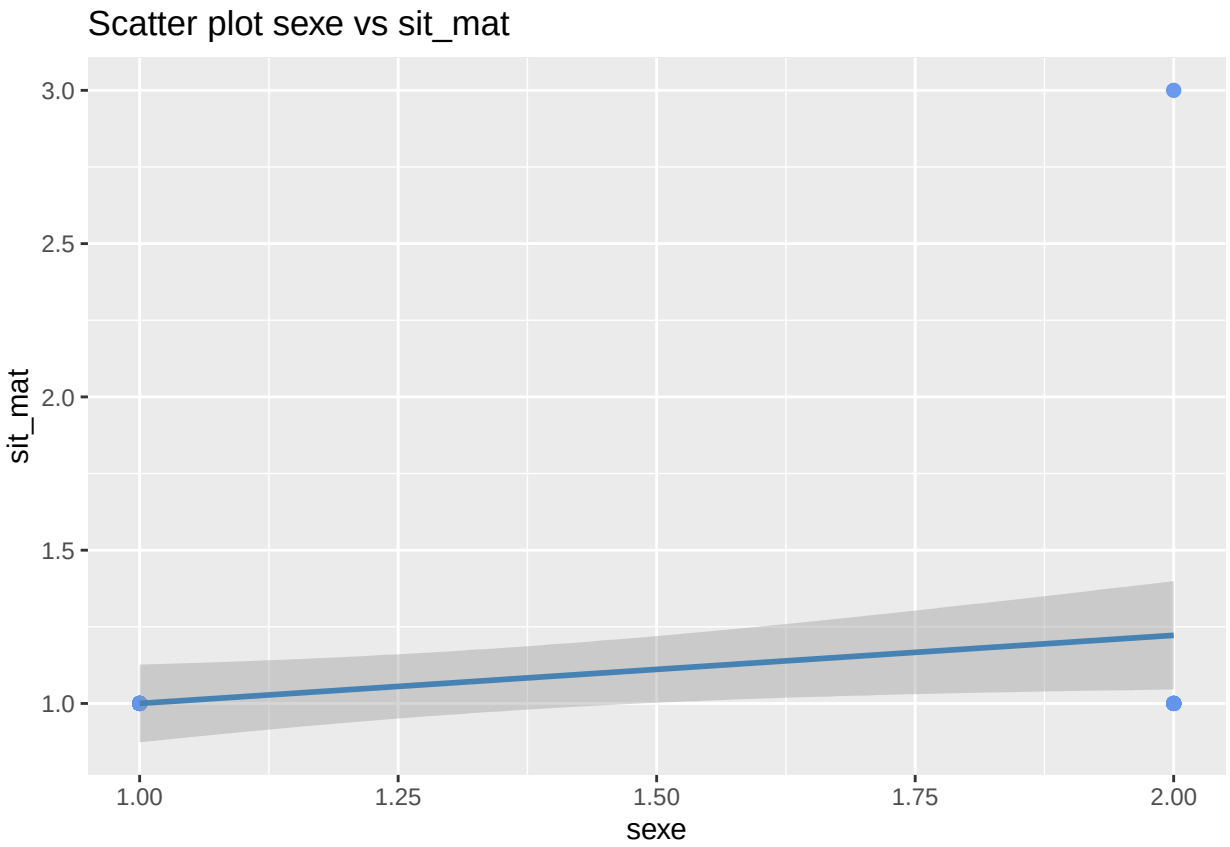
```
##
##      1  3
##   1 35  0
##   2 16  2

## $stacked
```


Stacked Bar chart of sexe and sit_mat
sit_mat
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

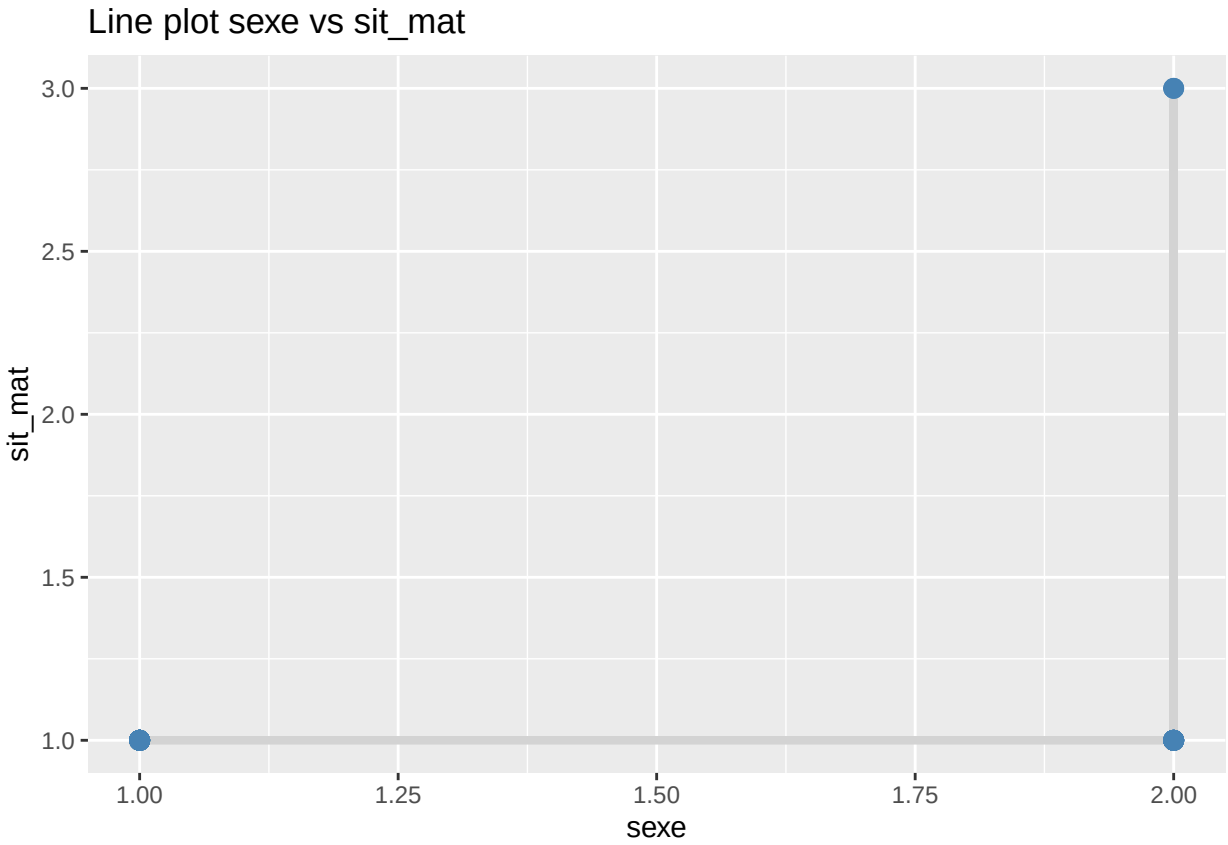
```
##
## $grouped
```


Grouped Bar chart of sexe and sit_mat
sit_mat
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

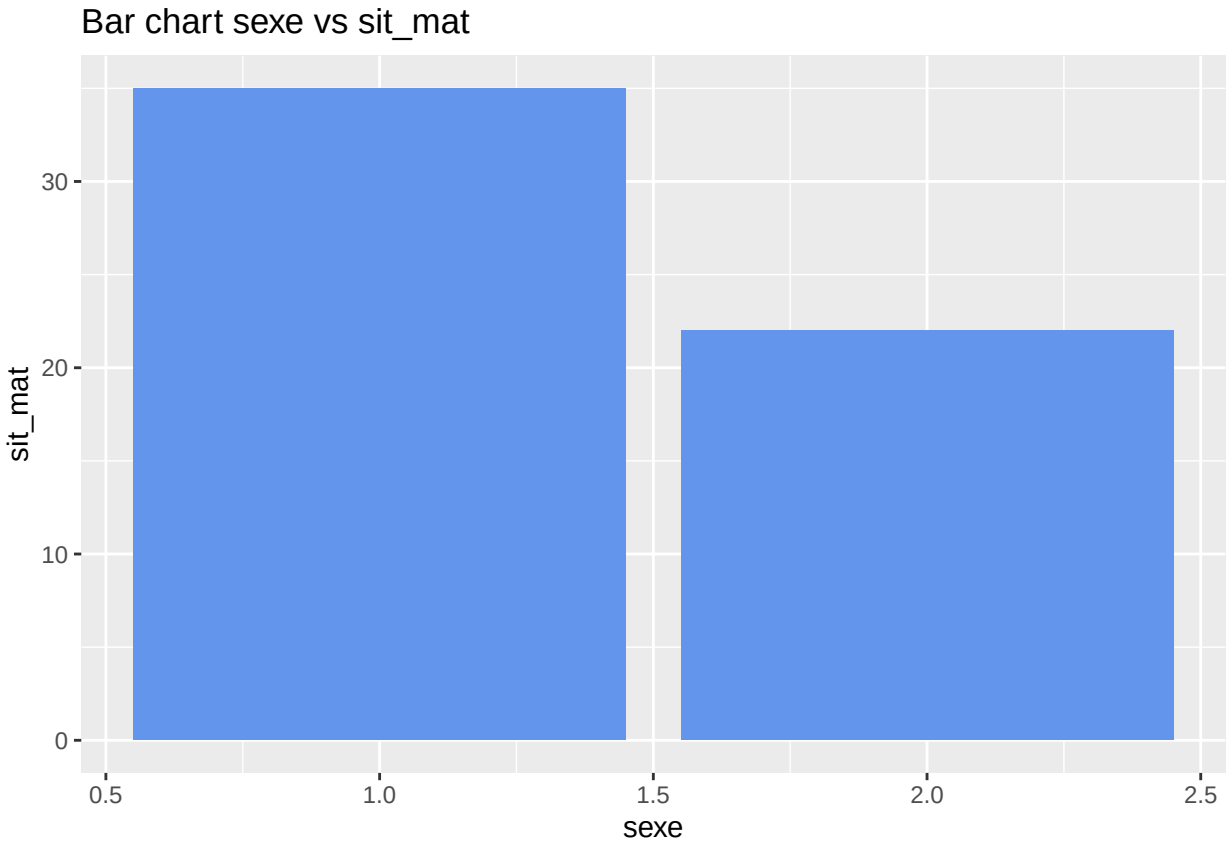
```
##
## $scatter
```


Scatter plot sexe vs sit_mat
3.0
2.5
2.0
1.5
1.0
sit_mat
1.00
1.25
1.50
1.75
2.00
sexe

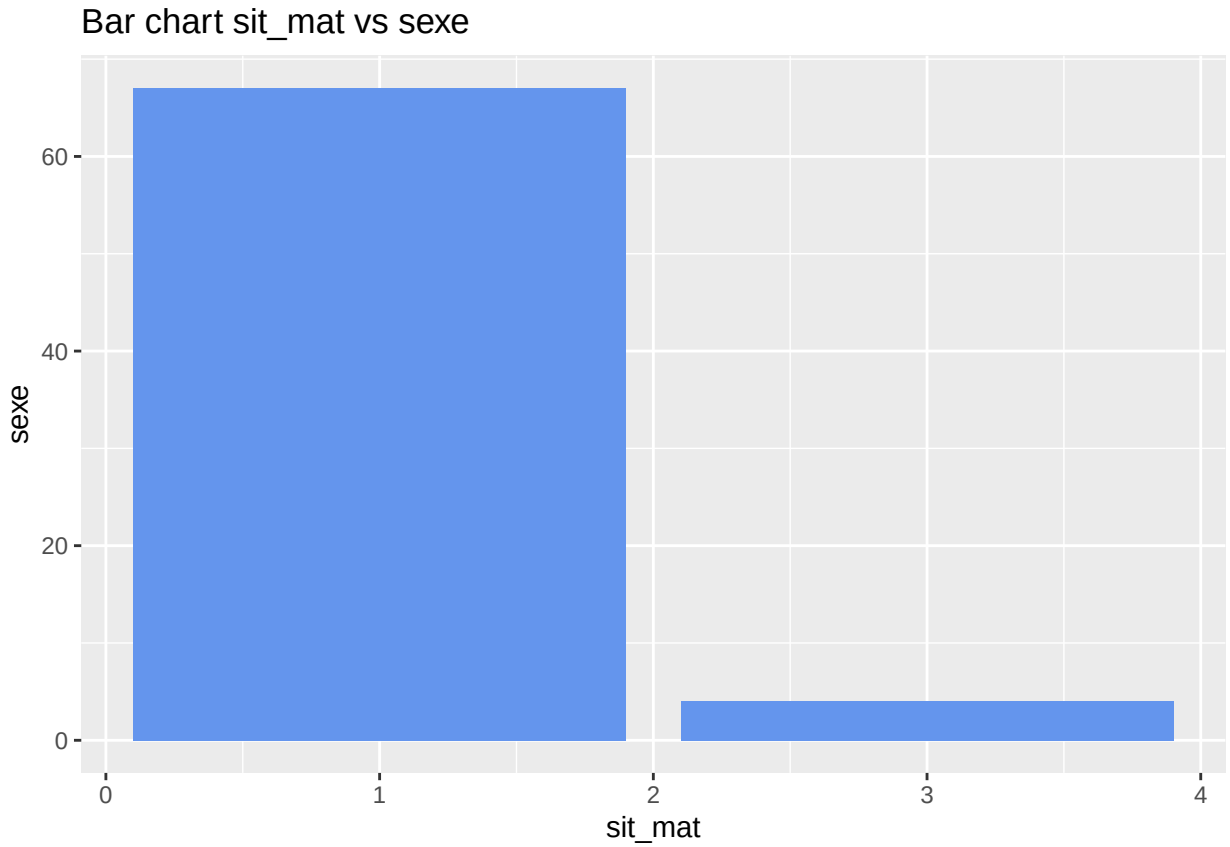
```
##
## $line
```


Line plot sexe vs sit_mat
3.0
2.5
2.0
1.5
1.0
sit_mat
1.00
1.25
1.50
1.75
2.00
sexe

```
##
## $bar_x
```




```
##
## $bar_y
```

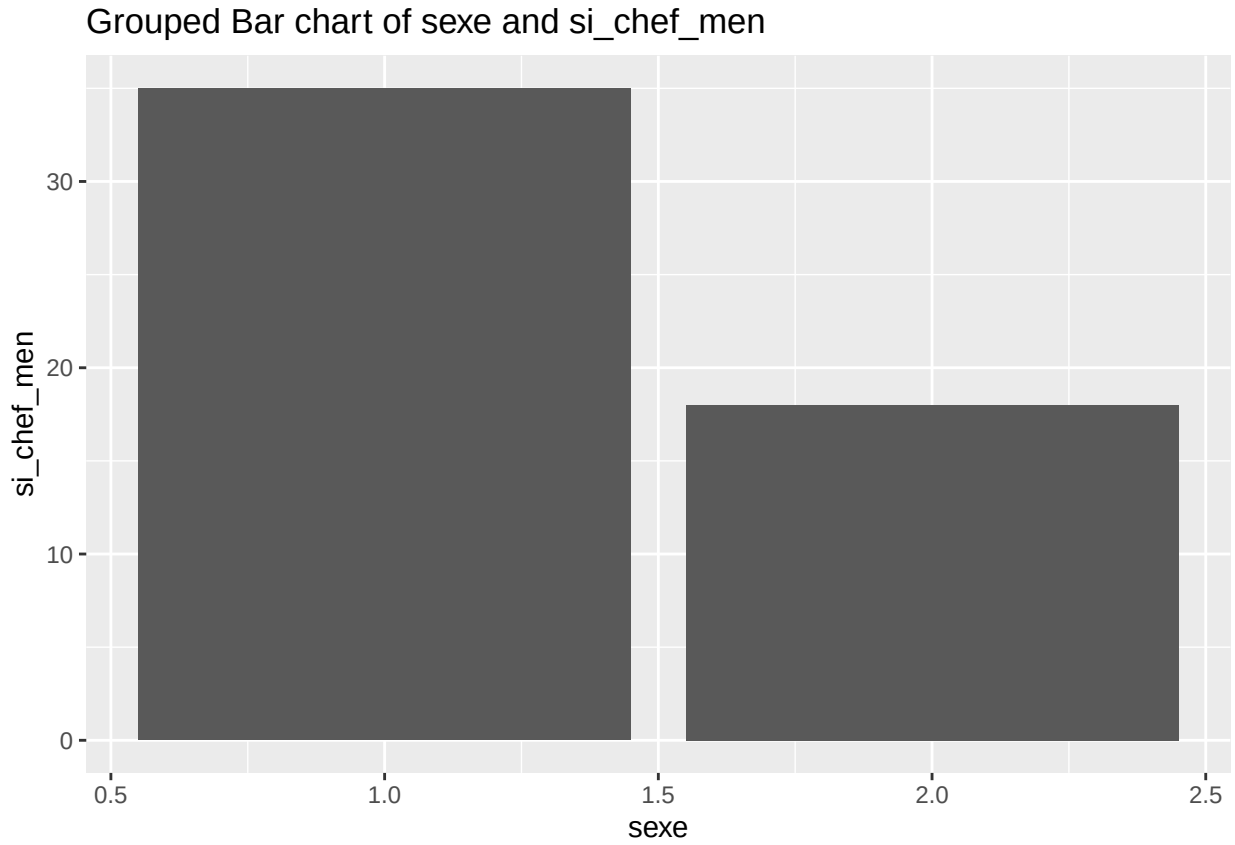



```
bivarie(base_tp2, "sexe", "si_chef_men", plot = TRUE)
```

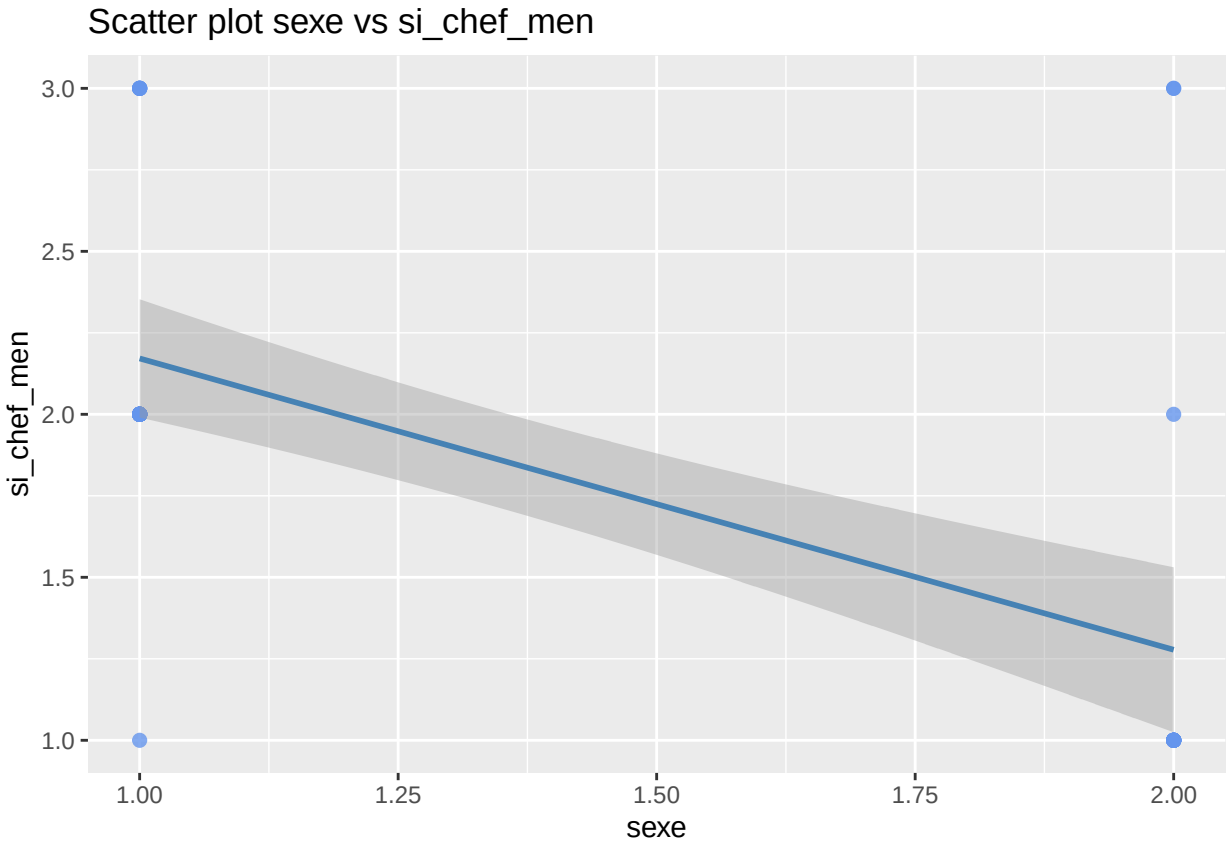
```
##
##       1  2  3
##    1  1 27  7
##    2 15  1  2

## $stacked
```


Stacked Bar chart of sexe and si_chef_men
si_chef_men
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

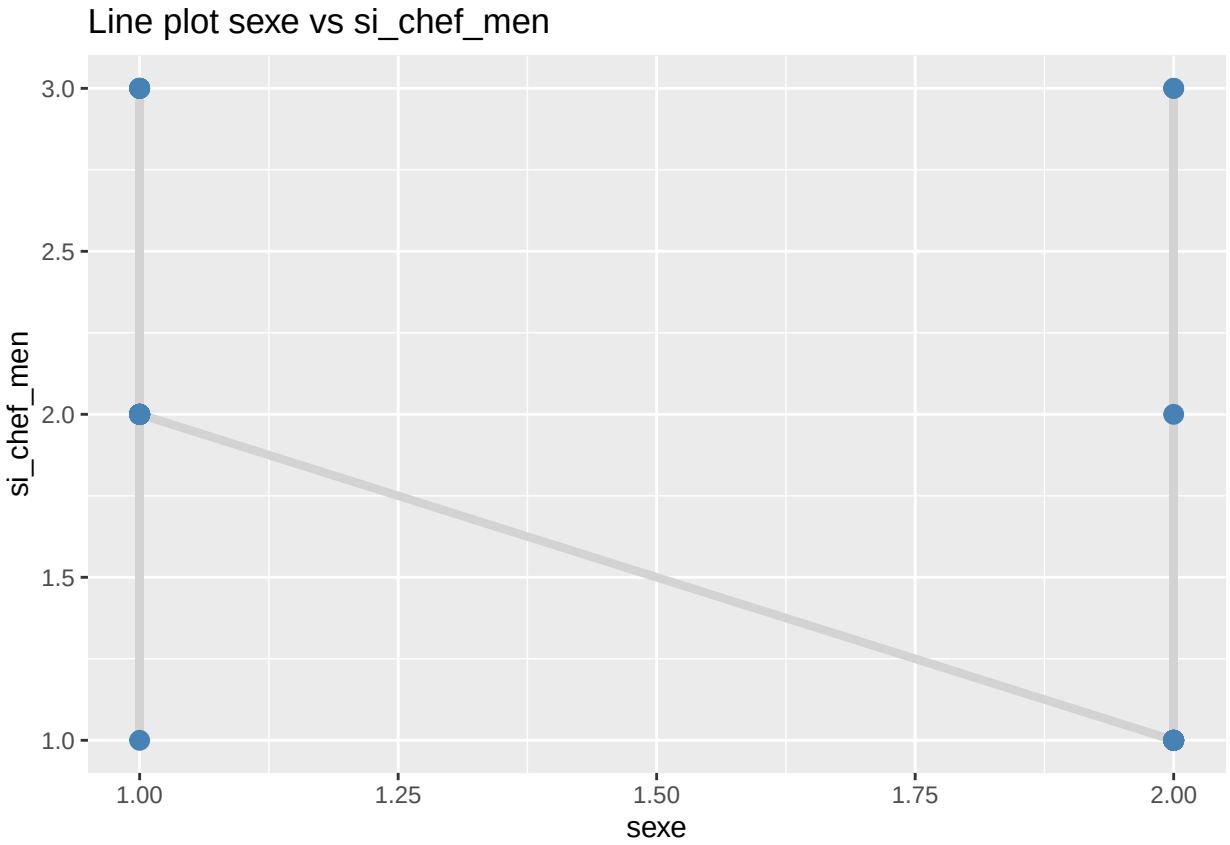
```
##
## $grouped
```


Grouped Bar chart of sexe and si_chef_men
si_chef_men
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

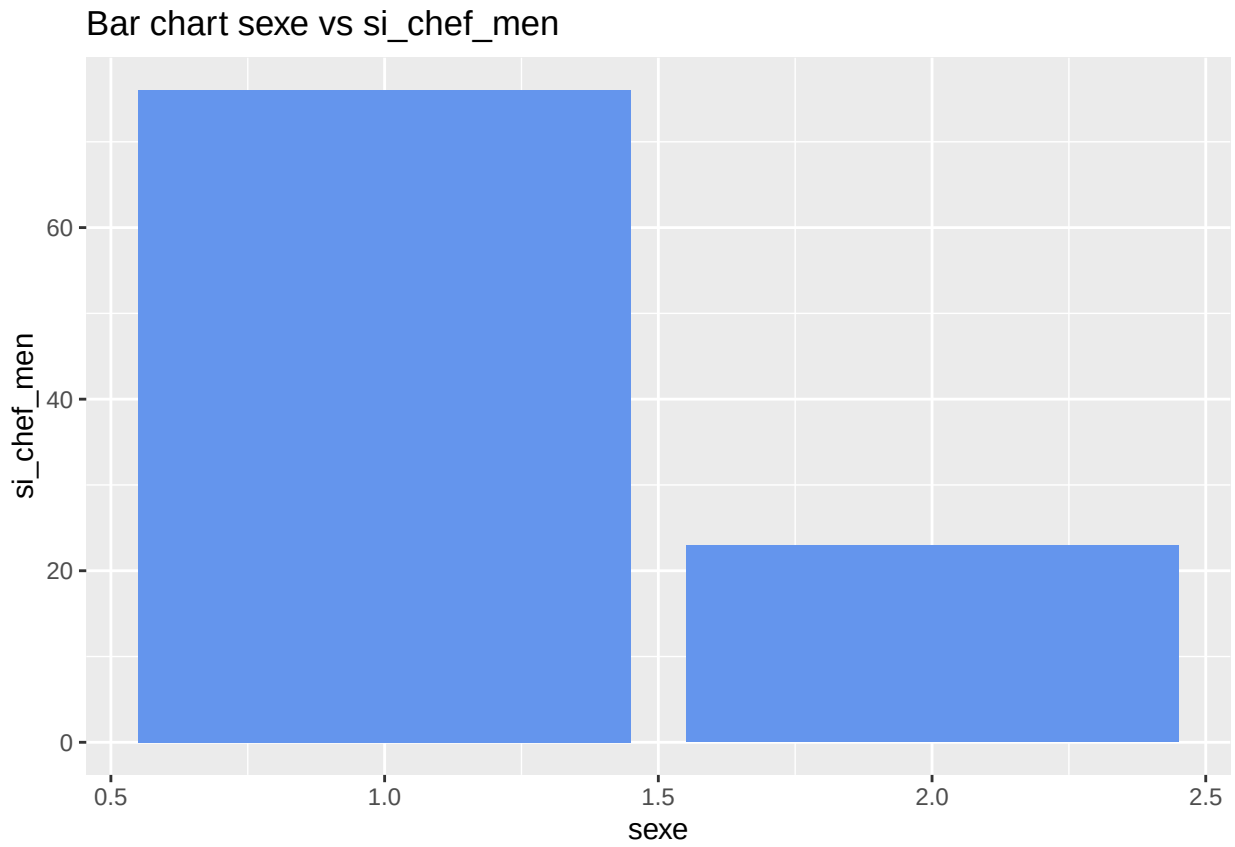
```
##
## $scatter
```


Scatter plot sexe vs si_chef_men
si_chef_men
3.0
2.5
2.0
1.5
1.0
1.00
1.25
1.50
1.75
2.00
sexe

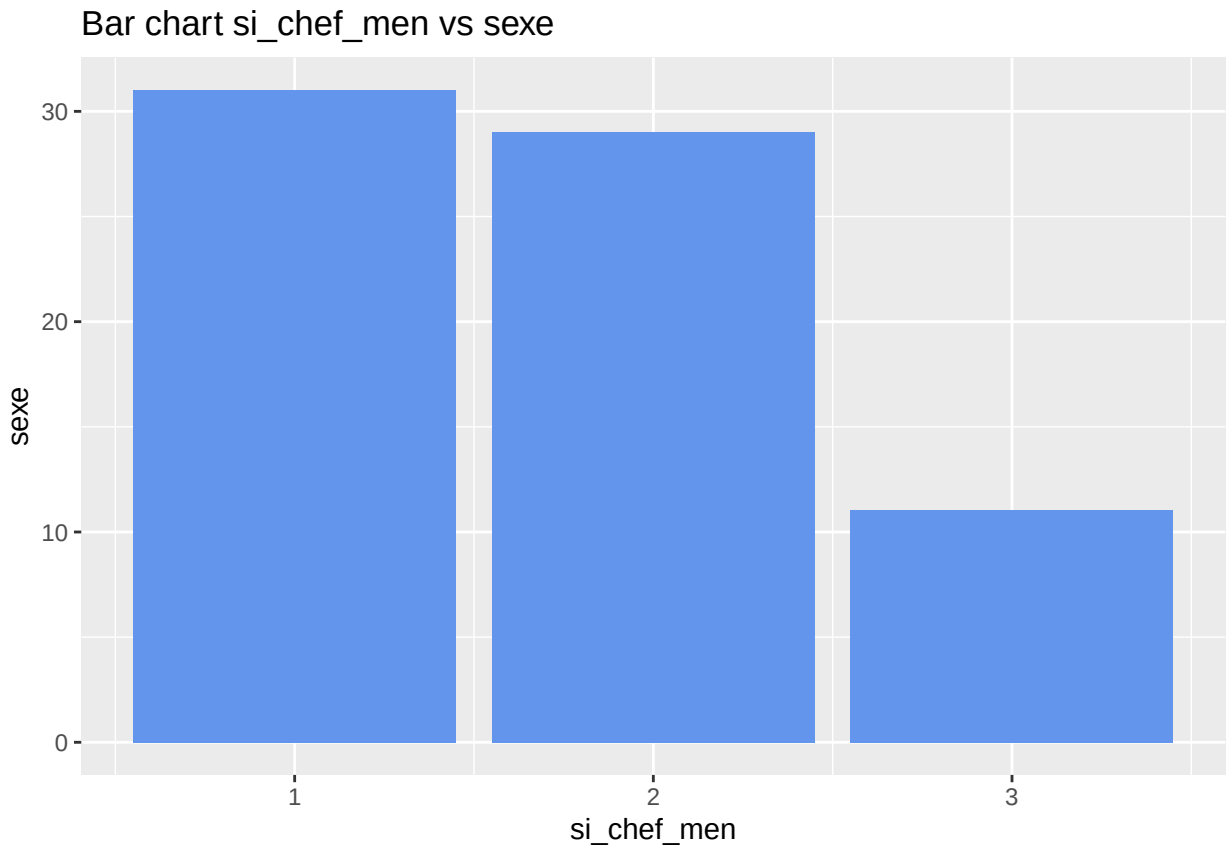
```
##
## $line
```


Line plot sexe vs si_chef_men
3.0
2.5
2.0
1.5
1.0
si_chef_men
1.00
1.25
1.50
1.75
2.00
sexe

```
##
## $bar_x
```

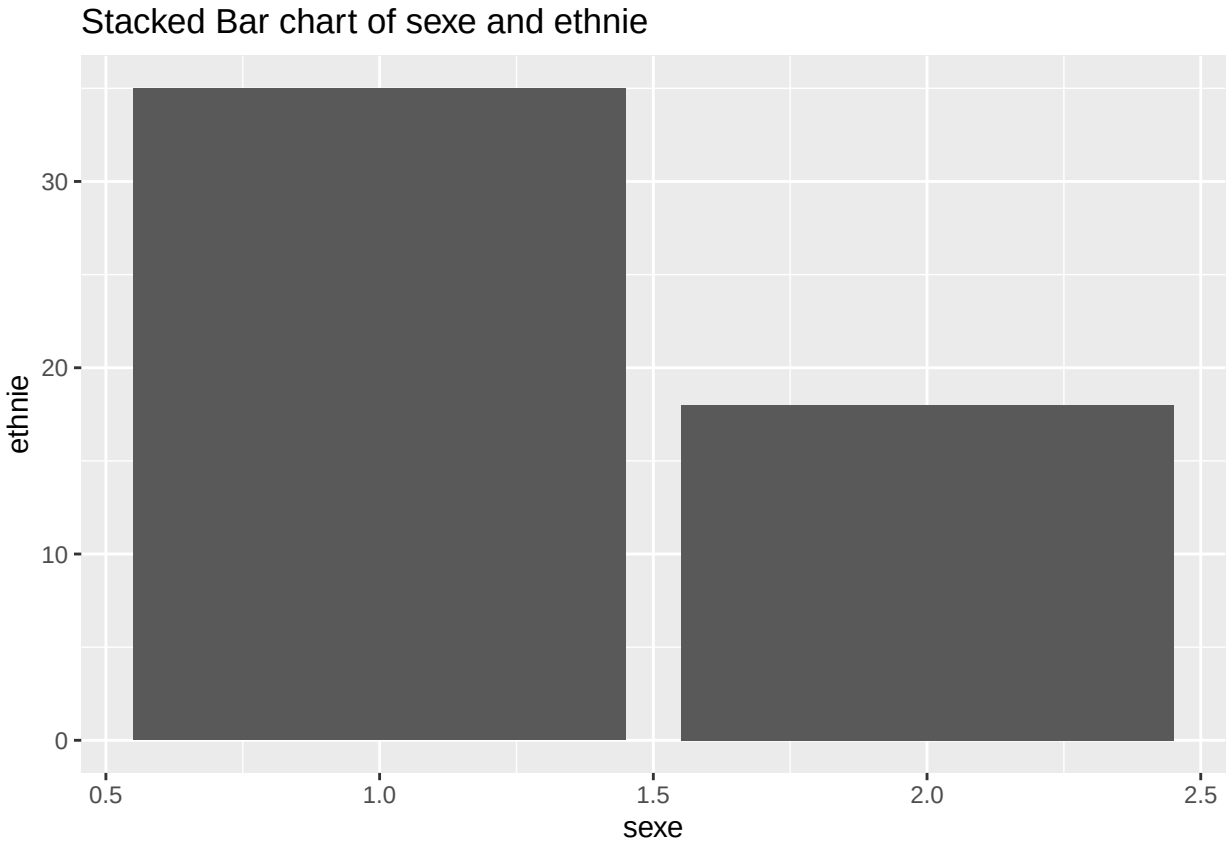

Bar chart sexe vs si_chef_men
si_chef_men
60
40
20
0
0.5
1.0
1.5
2.0
2.5
sexe

```
##
## $bar_y
```

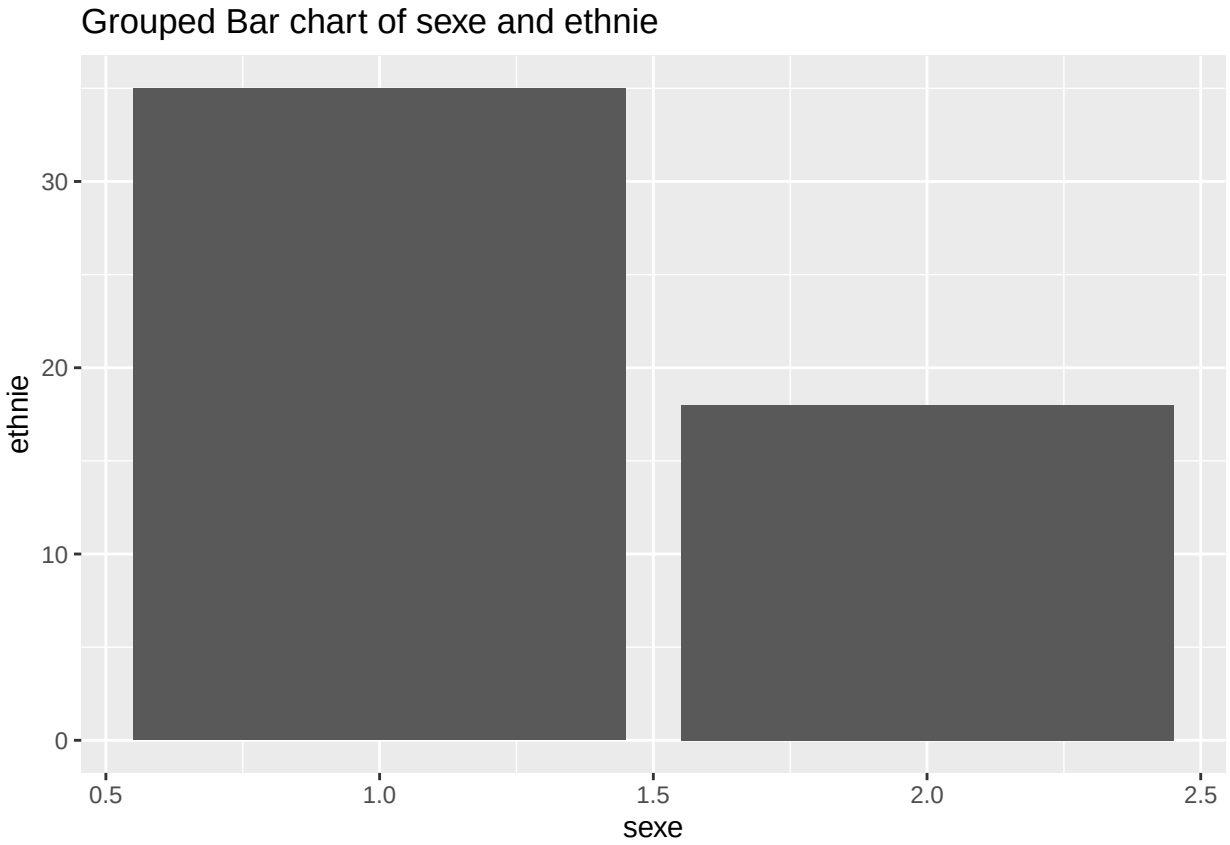



```
bivarie(base_tp2, "sexe", "ethnie", plot = TRUE)
```

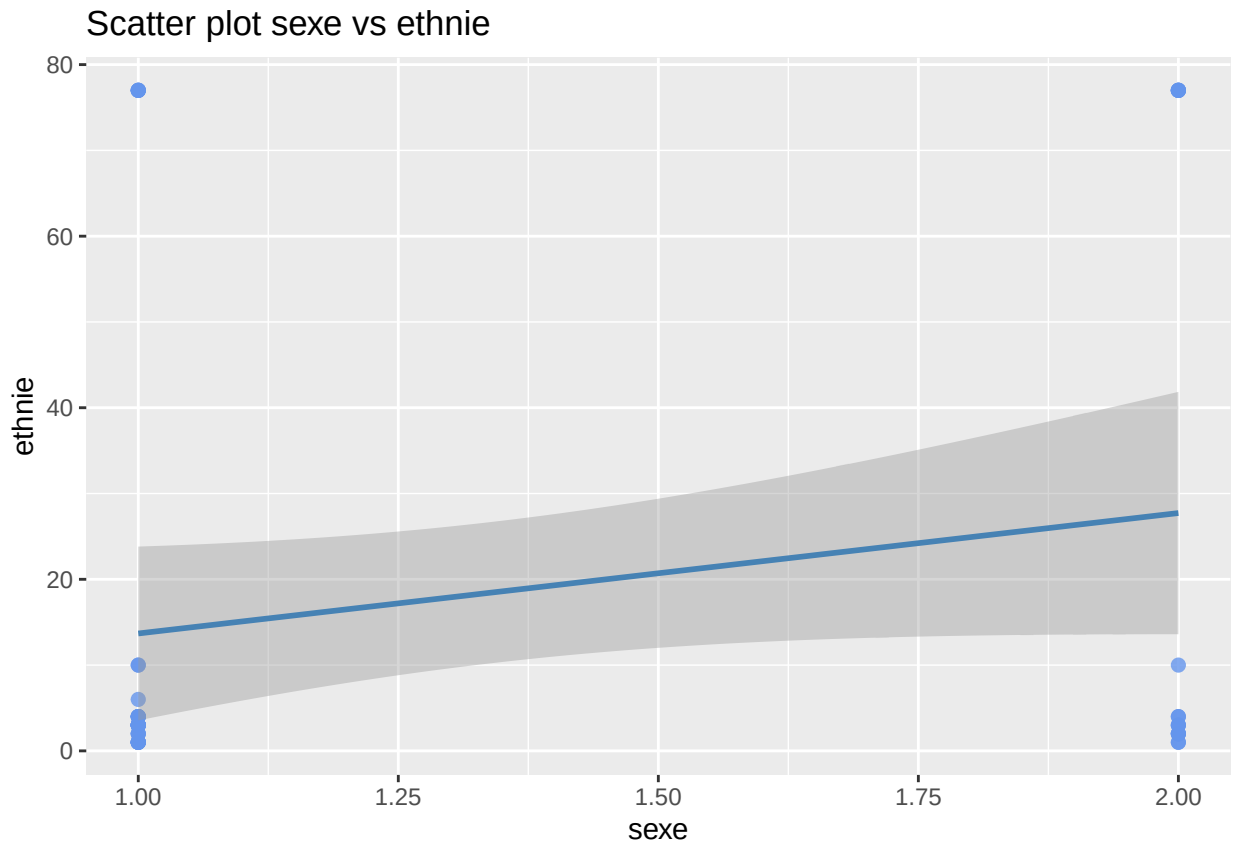
```
##
##      1  2  3  4  6 10 77
##   1  8  3 10  6  1  2  5
##   2  2  4  3  2  0  1  6

## $stacked
```


Stacked Bar chart of sexe and ethnie
ethnie
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

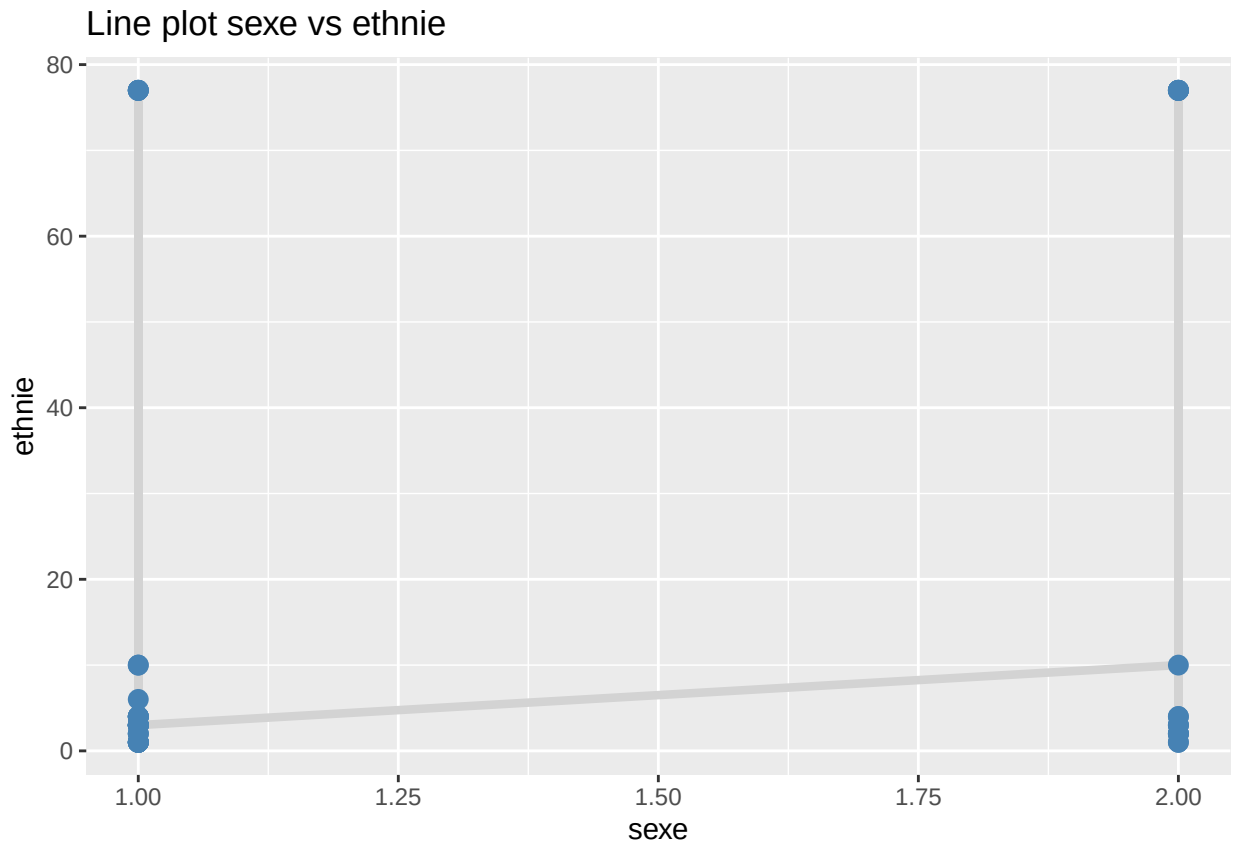
```
##
## $grouped
```


Grouped Bar chart of sexe and ethnie
ethnie
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

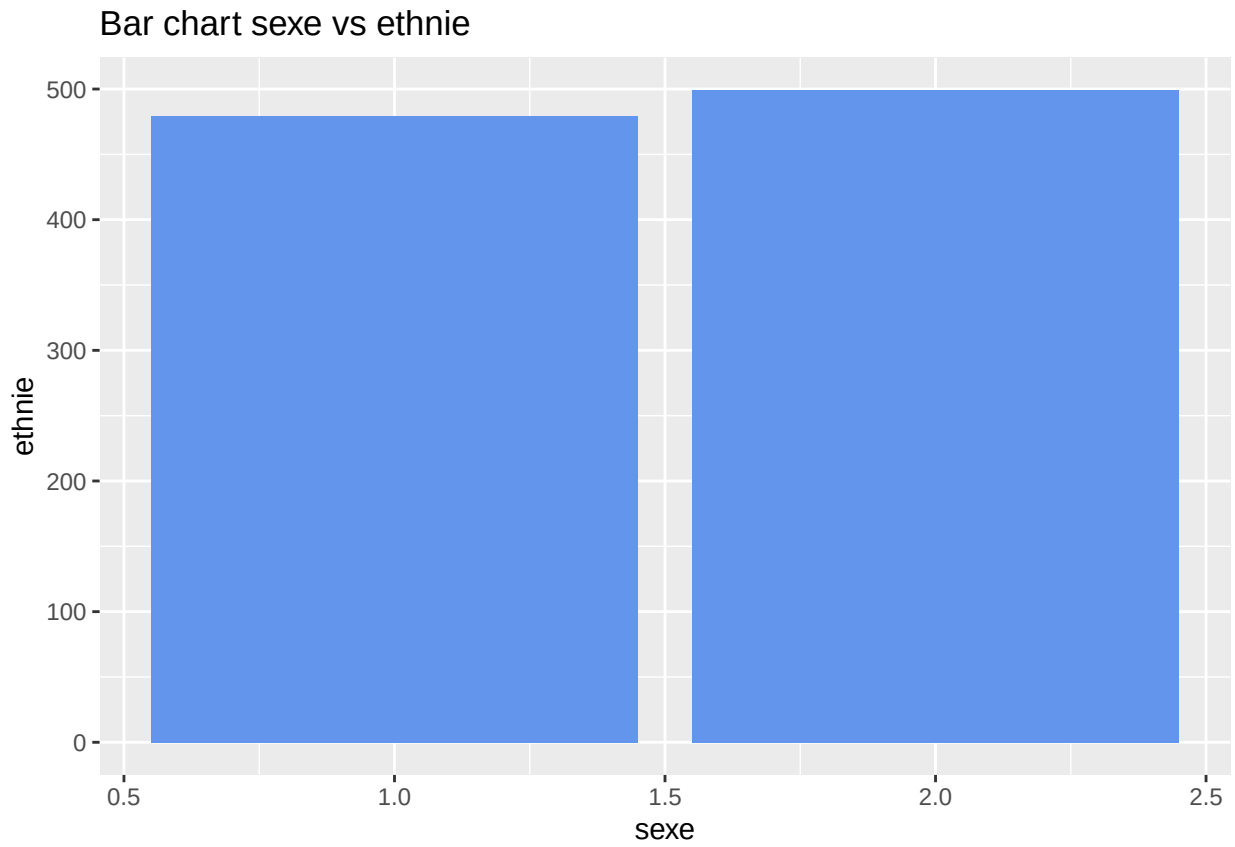
```
##
## $scatter
```


Scatter plot sexe vs ethnie
80
60
40
20
0
ethnie
1.00
1.25
1.50
1.75
2.00
sexe

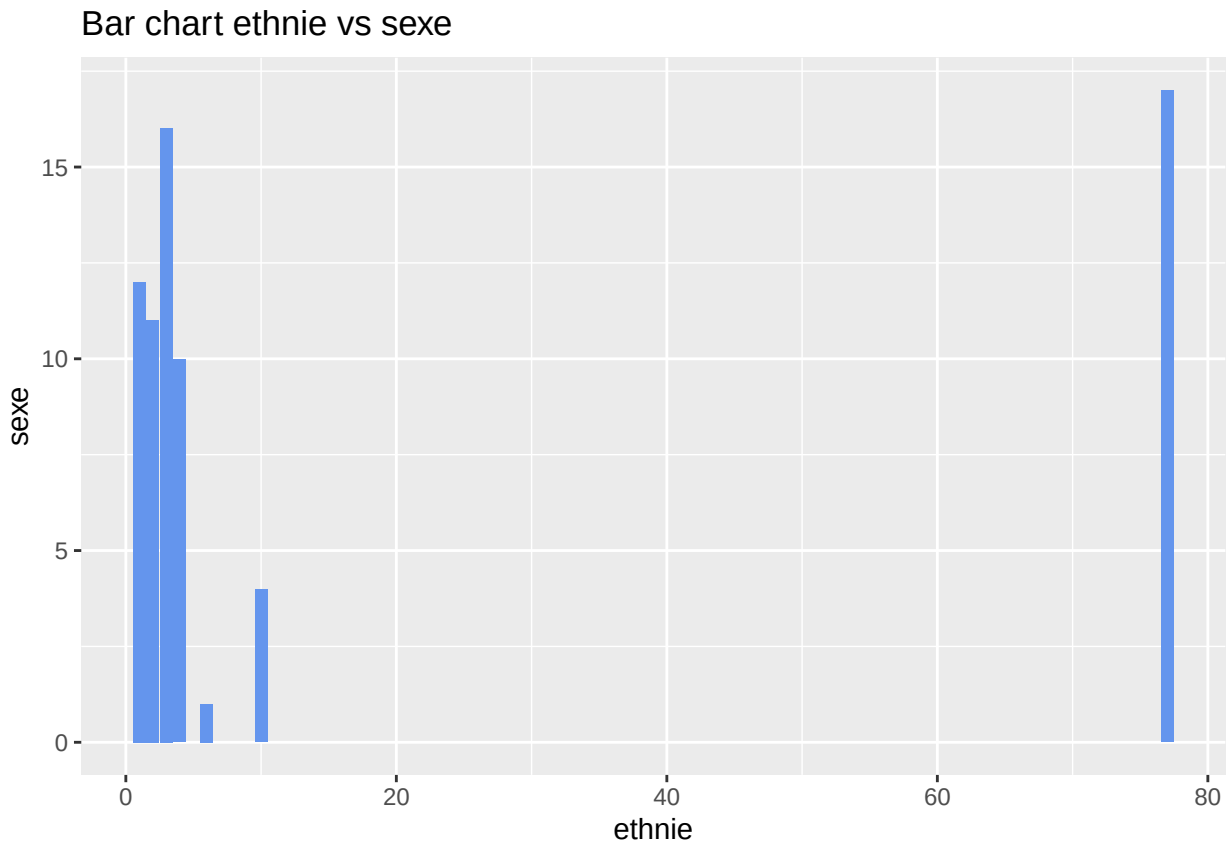
```
##
## $line
```


Line plot sexe vs ethnie
80
60
40
20
0
ethnie
1.00
1.25
1.50
1.75
2.00
sexe

```
##
## $bar_x
```

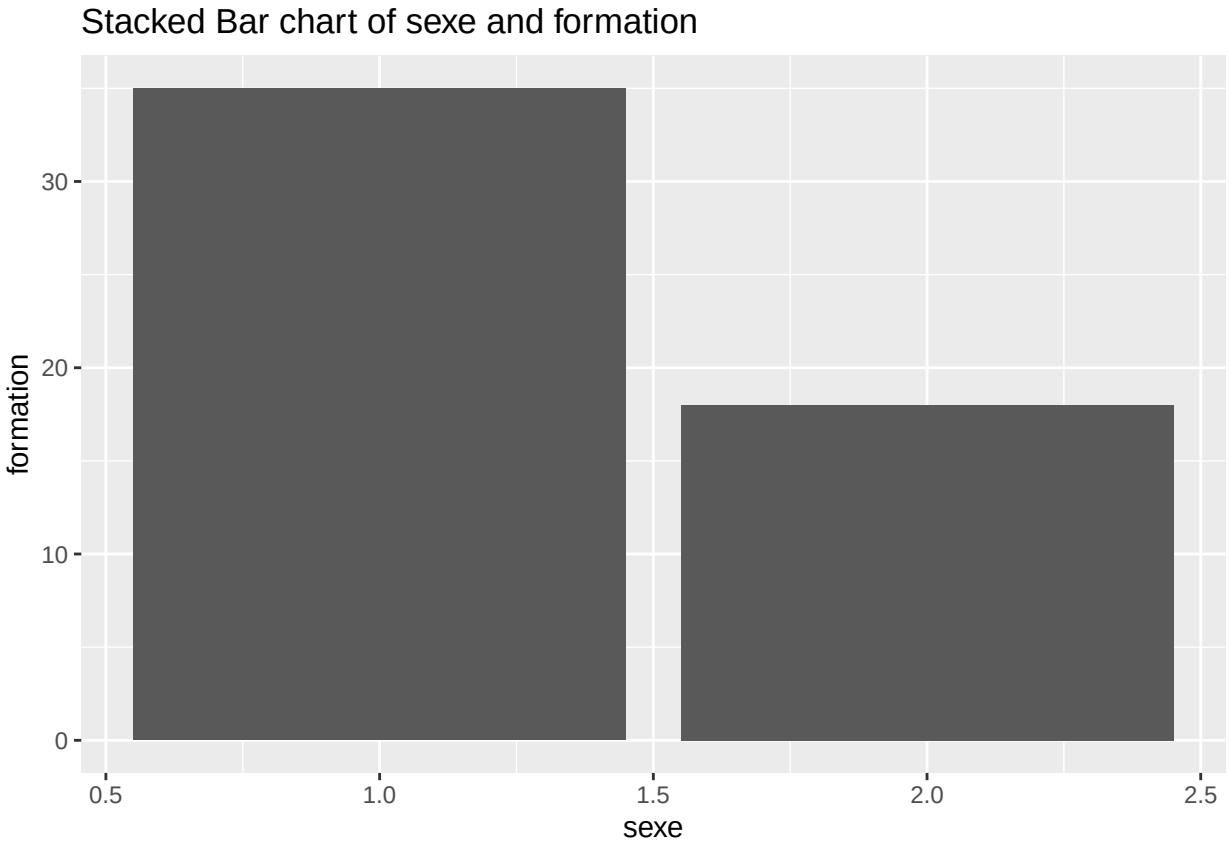

Bar chart sexe vs ethnie
ethnie
500
400
300
200
100
0
0.5
1.0
1.5
2.0
2.5
sexe

```
##
## $bar_y
```

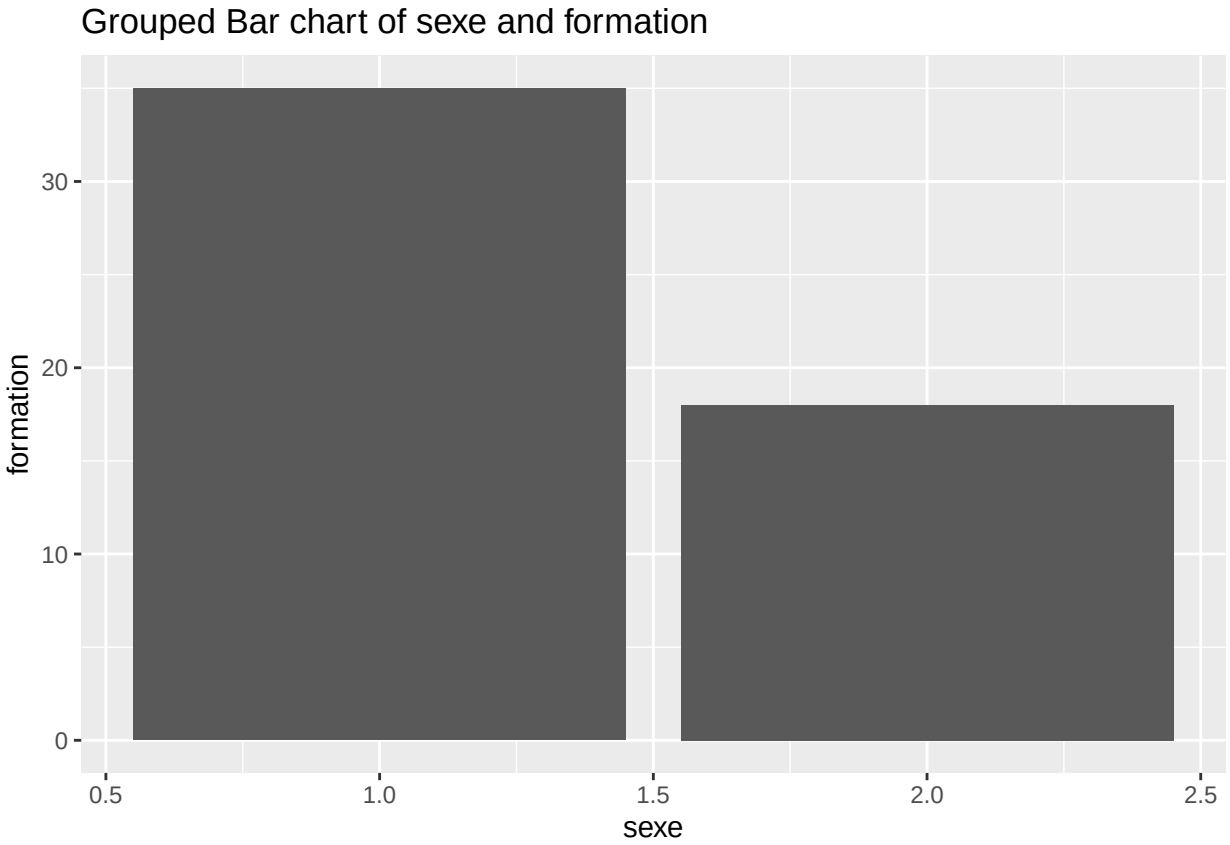



```
bivarie(base_tp2, "sexe", "formation", plot = TRUE)
```

```
##
##      1  2  3  4  5 99
##   1 17  8  5  2  2  1
##   2 13  2  2  1  0  0

## $stacked
```


Stacked Bar chart of sexe and formation
formation
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

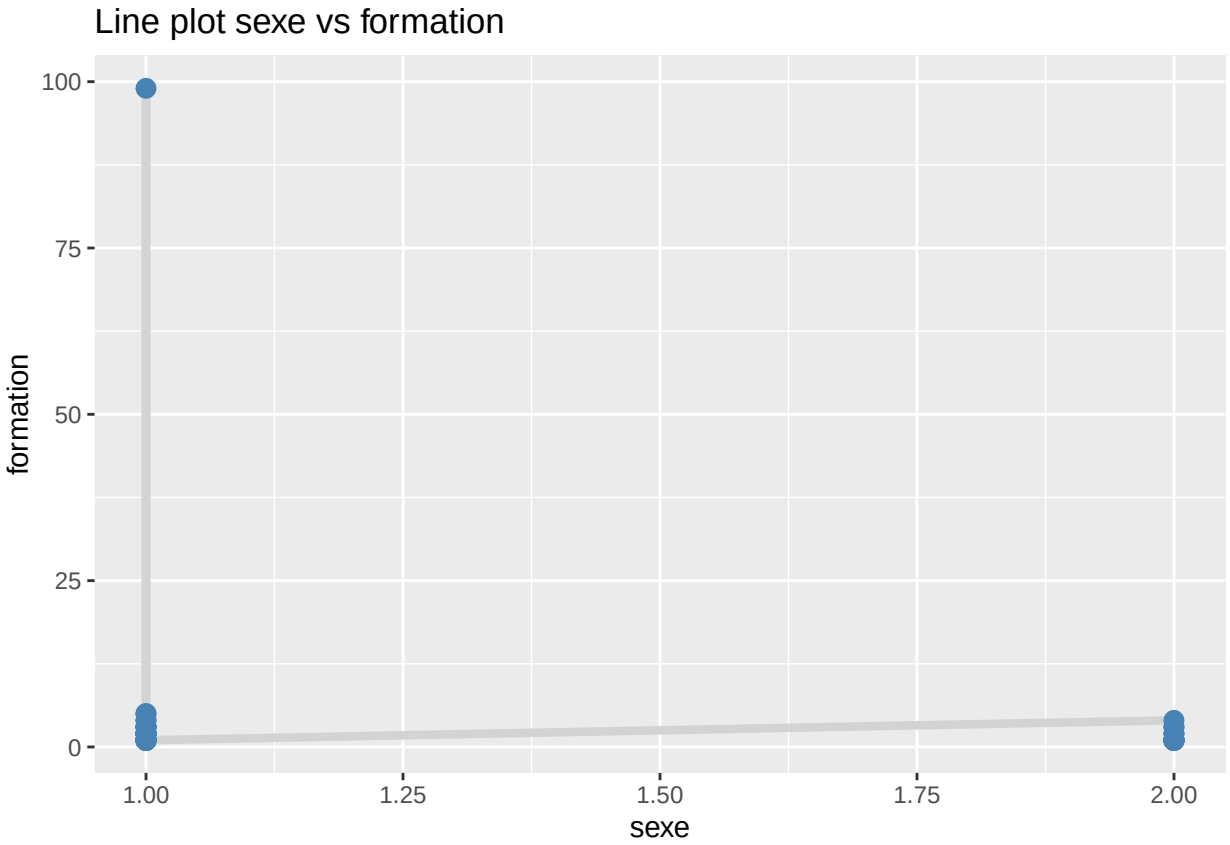
```
##
## $grouped
```


Grouped Bar chart of sexe and formation
formation
30
20
10
0
0.5
1.0
1.5
2.0
2.5
sexe

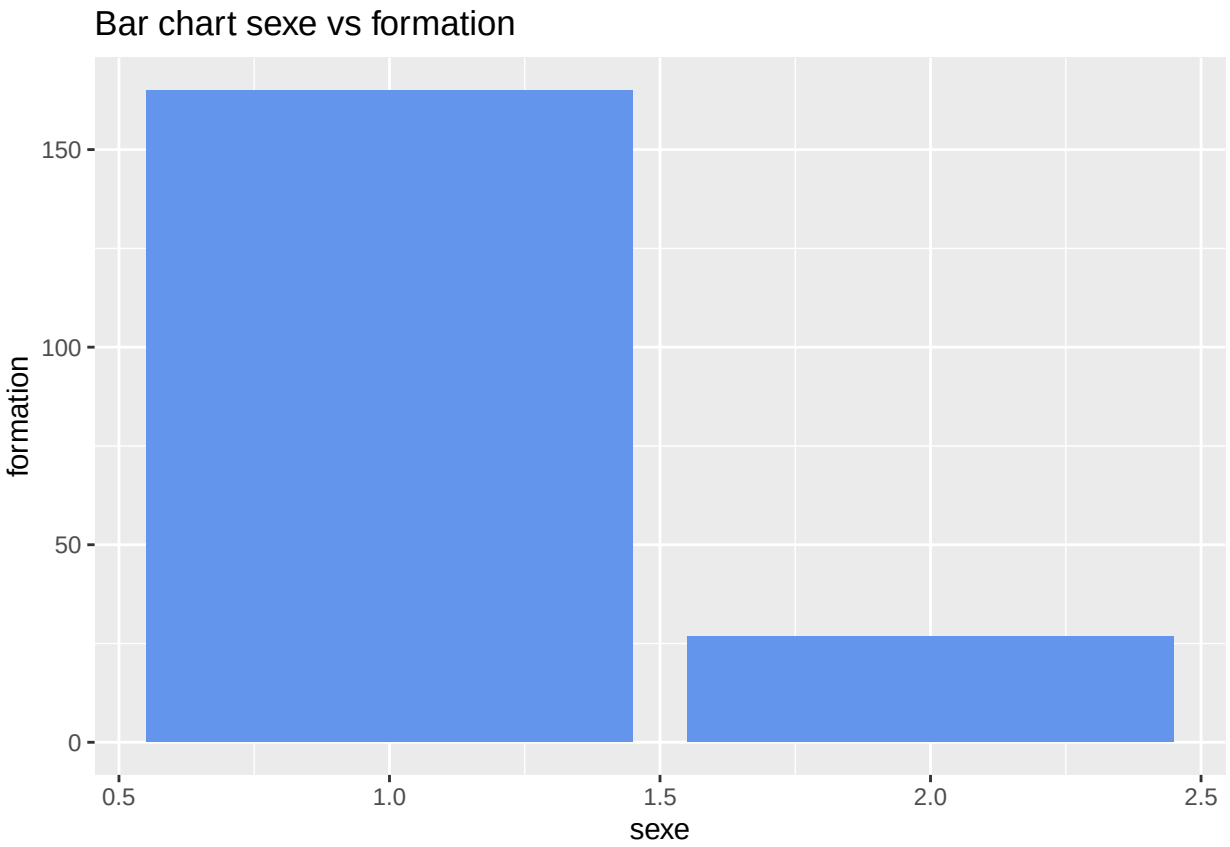
```
##
## $scatter
```


Scatter plot sexe vs formation
100
75
50
25
0
formation
1.00
1.25
1.50
1.75
2.00
sexe

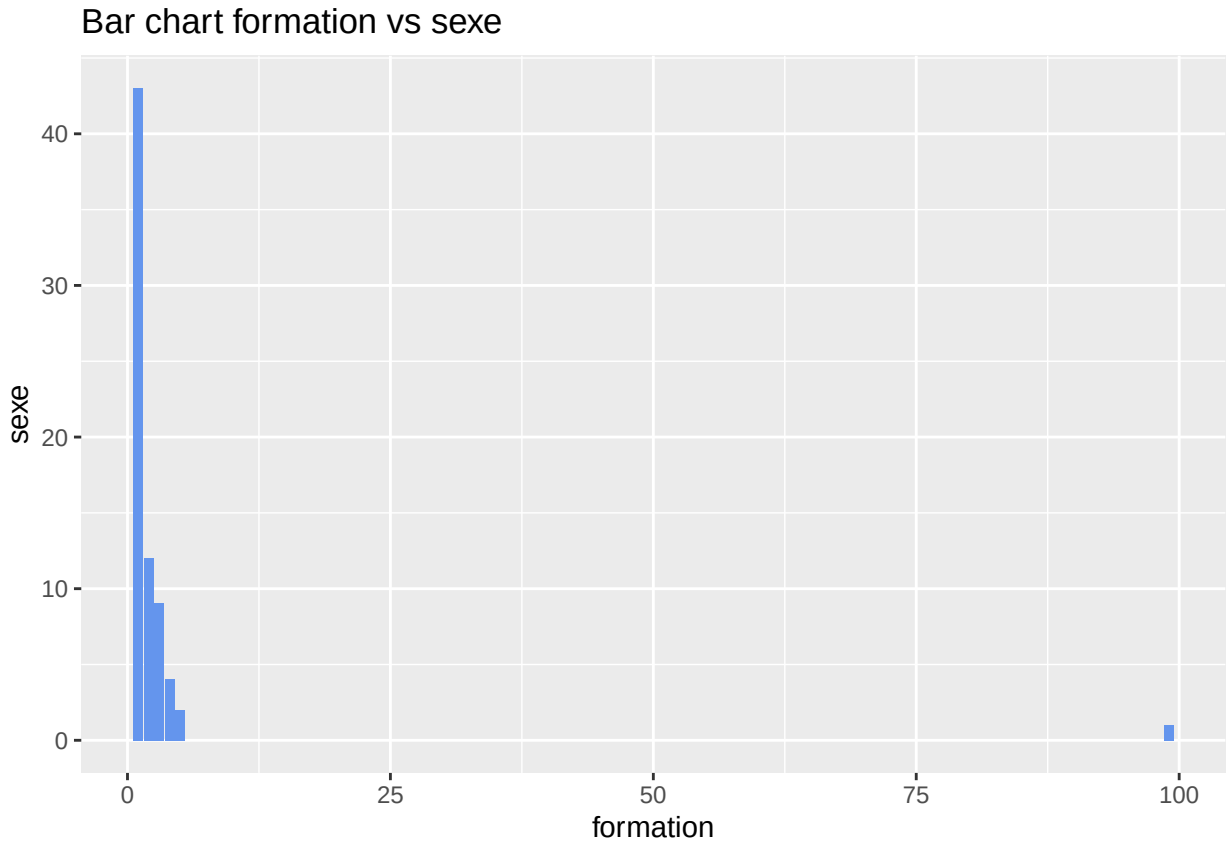
```
##
## $line
```


Line plot sexe vs formation
100
75
50
25
0
formation
1.00
1.25
1.50
1.75
2.00
sexe

```
##
## $bar_x
```

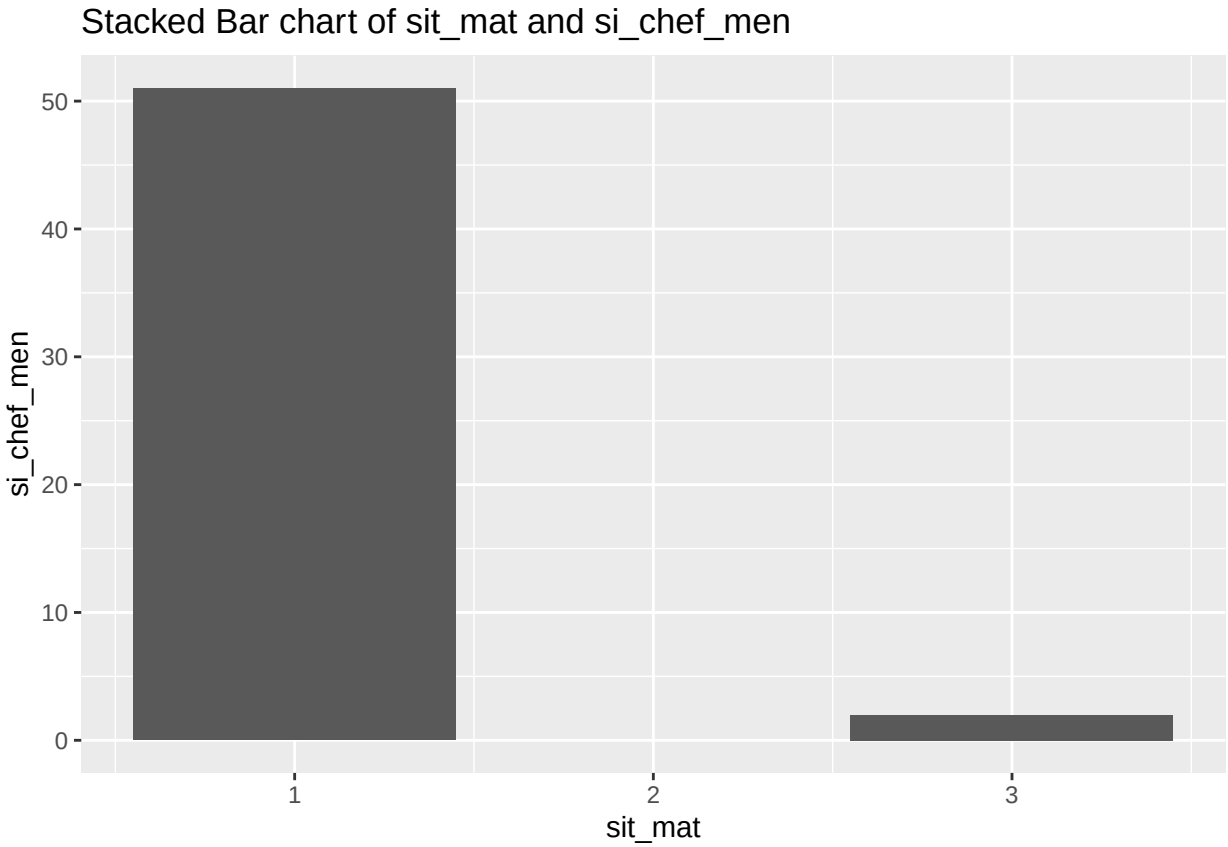

Bar chart sexe vs formation
150
100
50
0
formation
0.5
1.0
1.5
2.0
2.5
sexe

```
##
## $bar_y
```

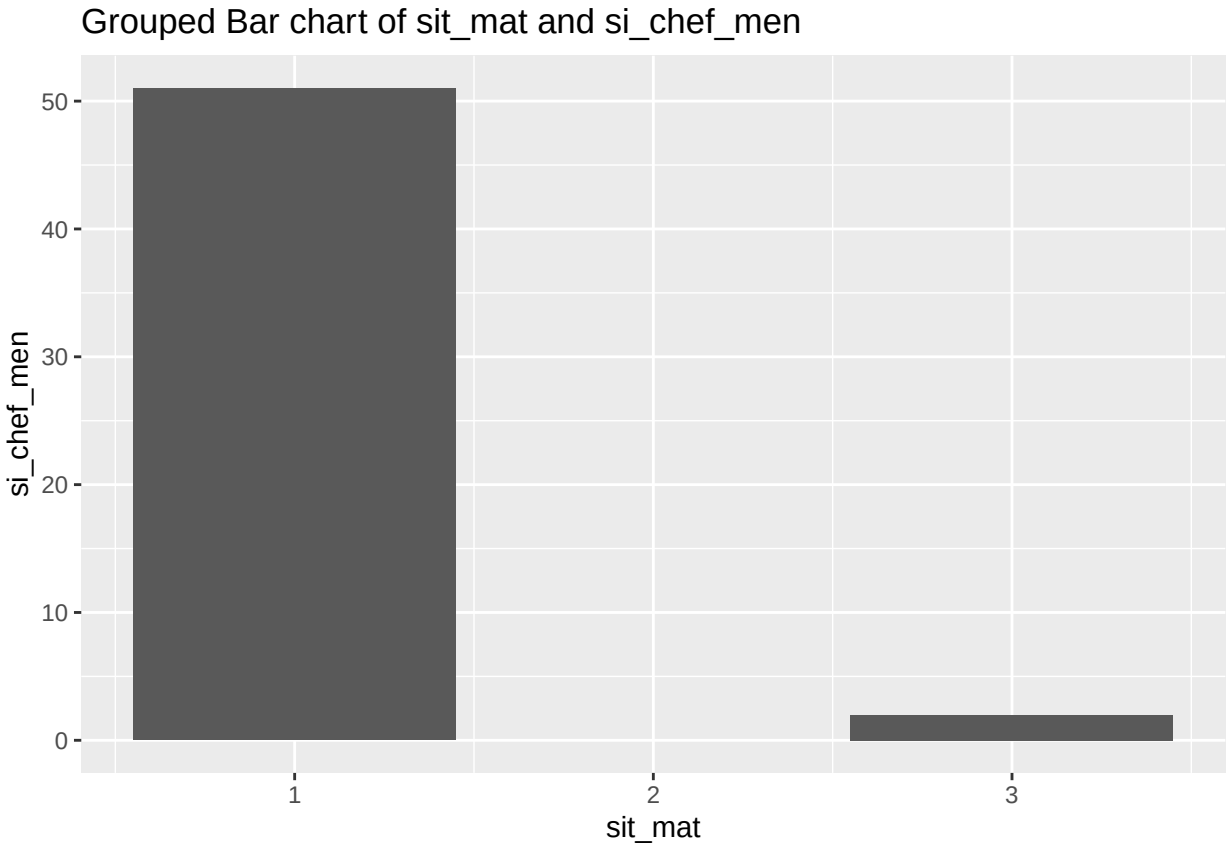



```
bivarie(base_tp2, "sit_mat", "si_chef_men", plot = TRUE)
```

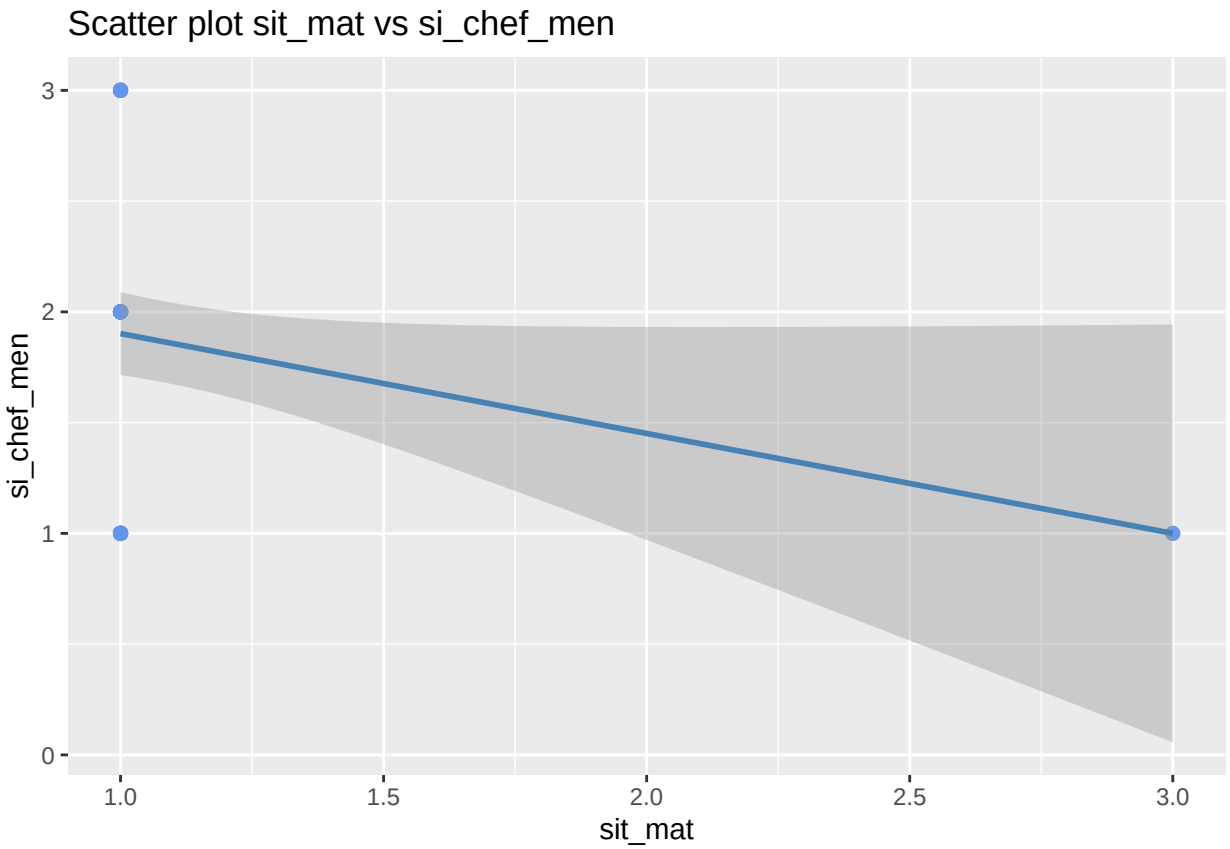
```
##
##      1  2  3
##   1 14 28  9
##   3  2  0  0

## $stacked
```


Stacked Bar chart of sit_mat and si_chef_men
si_chef_men
50
40
30
20
10
0
1
2
3
sit_mat

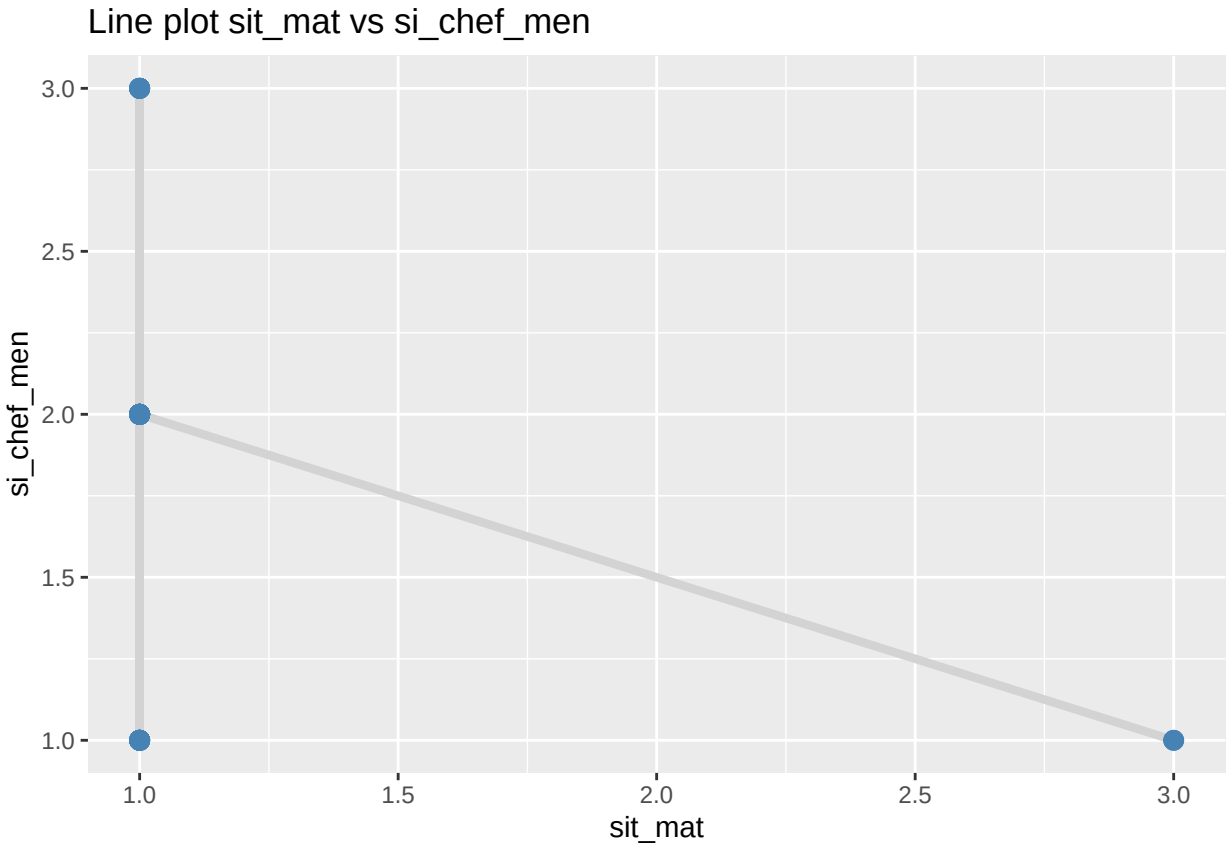
```
##
## $grouped
```


Grouped Bar chart of sit_mat and si_chef_men
si_chef_men
50
40
30
20
10
0
1
2
3
sit_mat

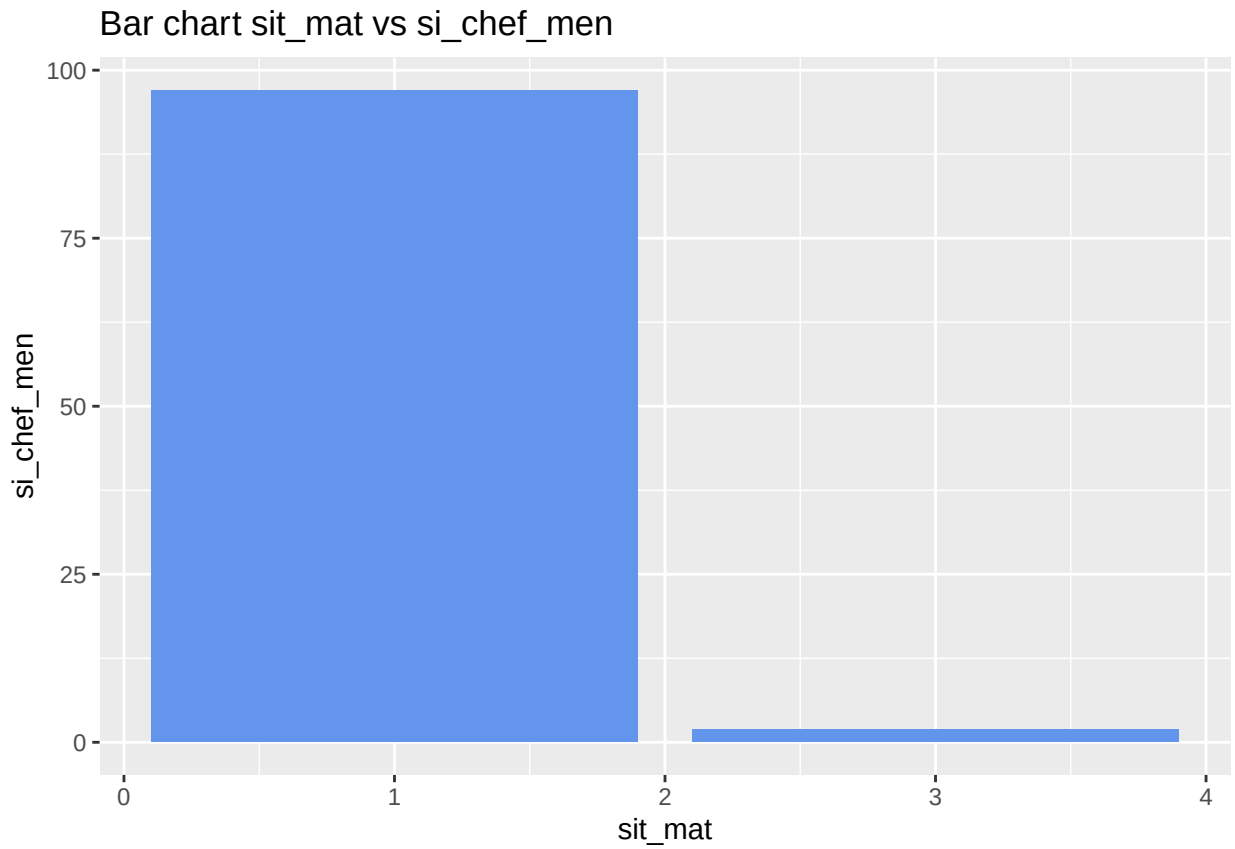
```
##
## $scatter
```


Scatter plot sit_mat vs si_chef_men
si_chef_men
sit_mat
3
2
1
0
1.0
1.5
2.0
2.5
3.0

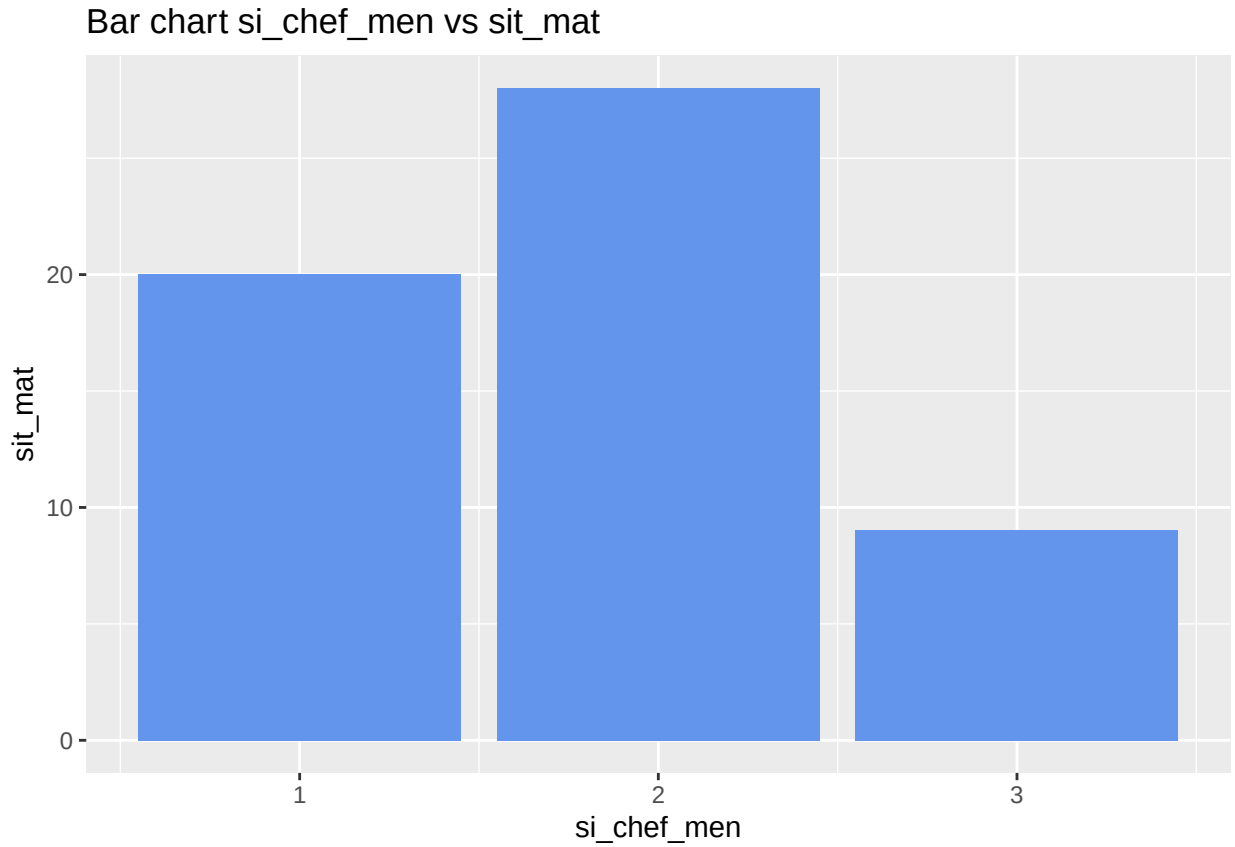
```
##
## $line
```


Line plot sit_mat vs si_chef_men
3.0
2.5
2.0
1.5
1.0
si_chef_men
1.0
1.5
2.0
2.5
3.0
sit_mat

```
##
## $bar_x
```

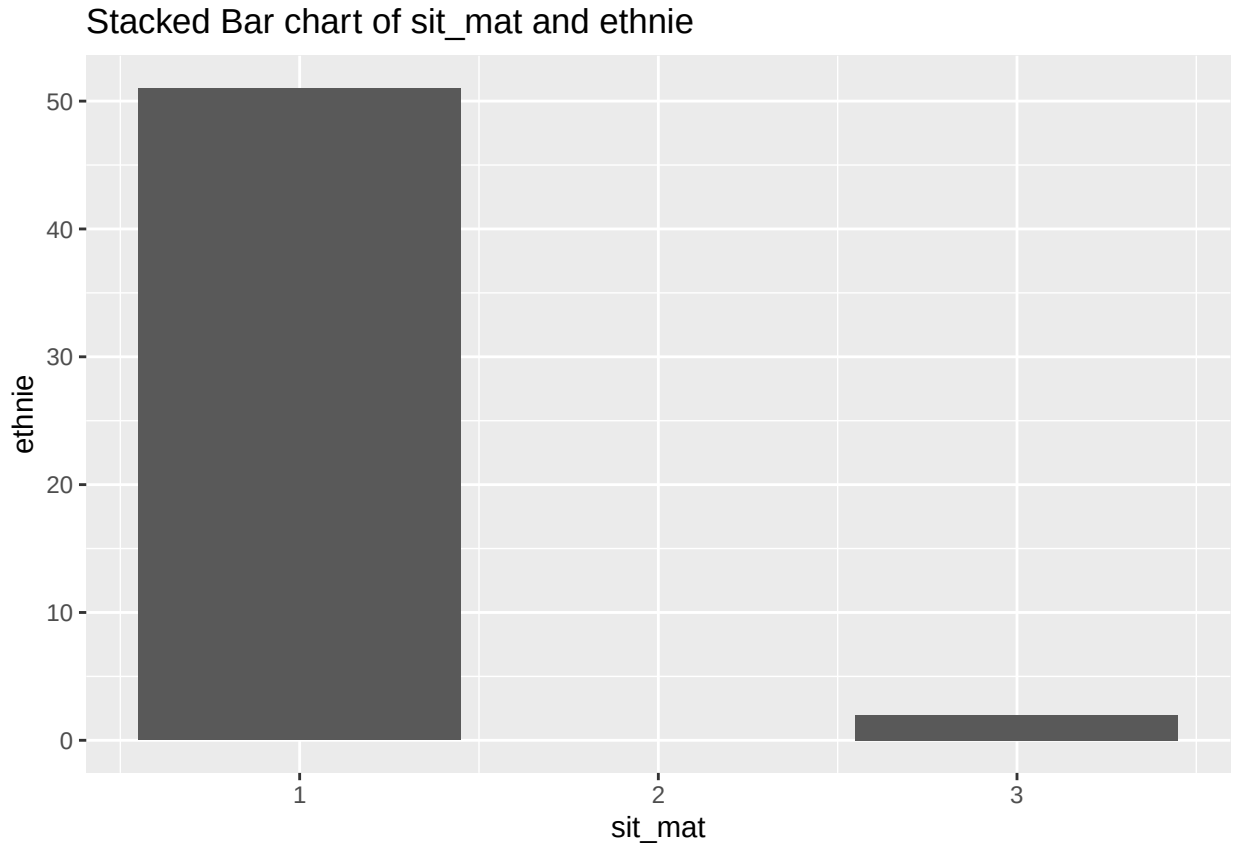

Bar chart sit_mat vs si_chef_men
100
75
50
25
0
si_chef_men
0
1
2
3
4
sit_mat

```
##
## $bar_y
```

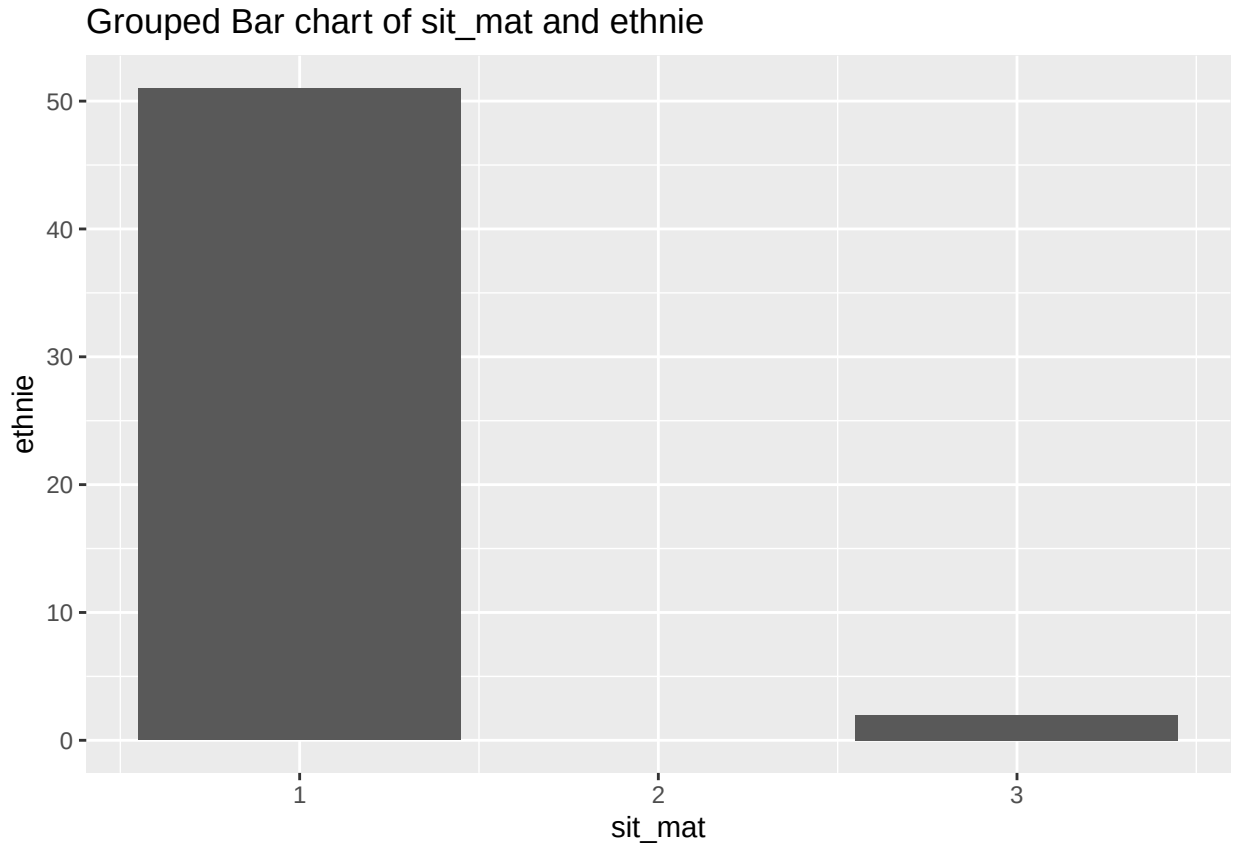



```
bivarie(base_tp2, "sit_mat", "ethnie", plot = TRUE)
```

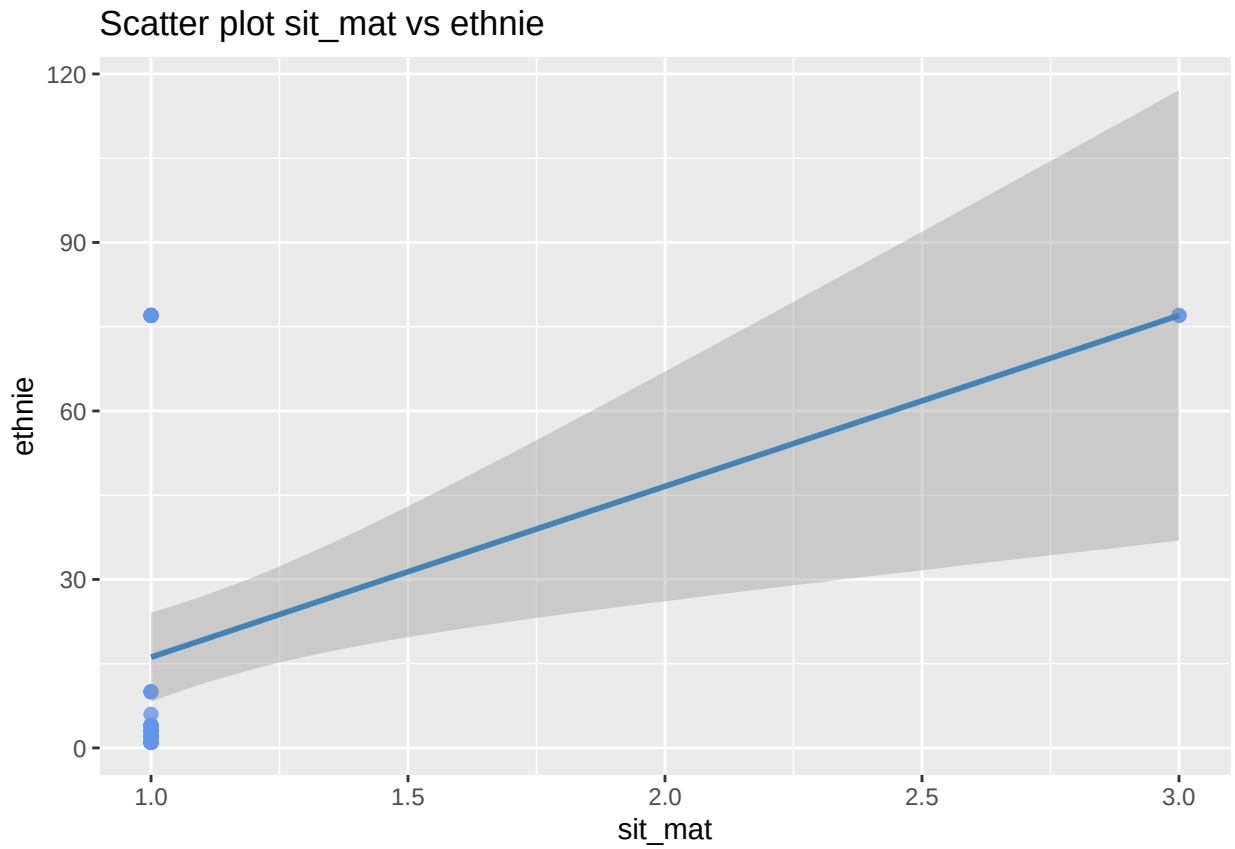
```
##
##       1  2  3  4  6 10 77
##   1 10  7 13  8  1  3  9
##   3  0  0  0  0  0  0  2

## $stacked
```


Stacked Bar chart of sit_mat and ethnie
ethnie
50
40
30
20
10
0
1
2
3
sit_mat

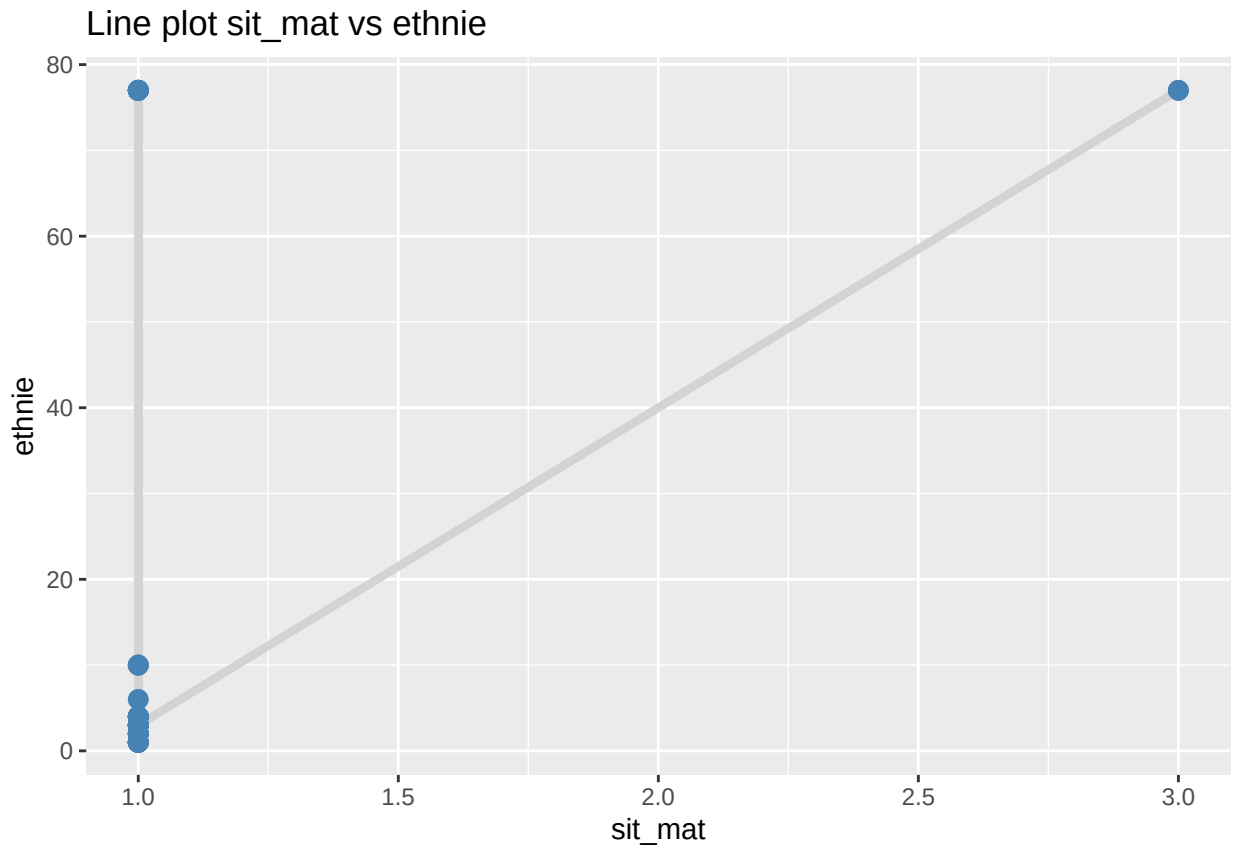
```
##
## $grouped
```


Grouped Bar chart of sit_mat and ethnie
ethnie
50
40
30
20
10
0
1
2
3
sit_mat

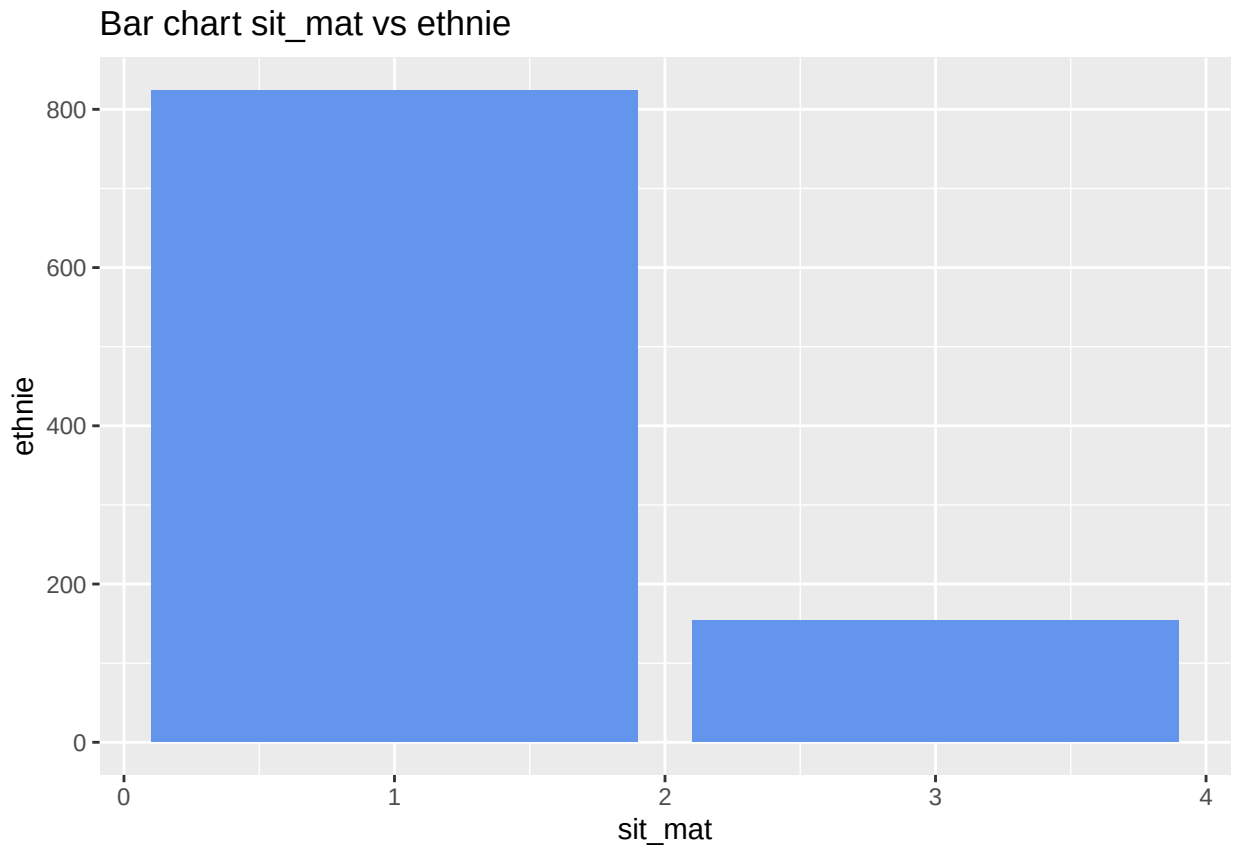
```
##
## $scatter
```


Scatter plot sit_mat vs ethnie
120
90
60
30
0
ethnie
1.0
1.5
2.0
2.5
3.0
sit_mat

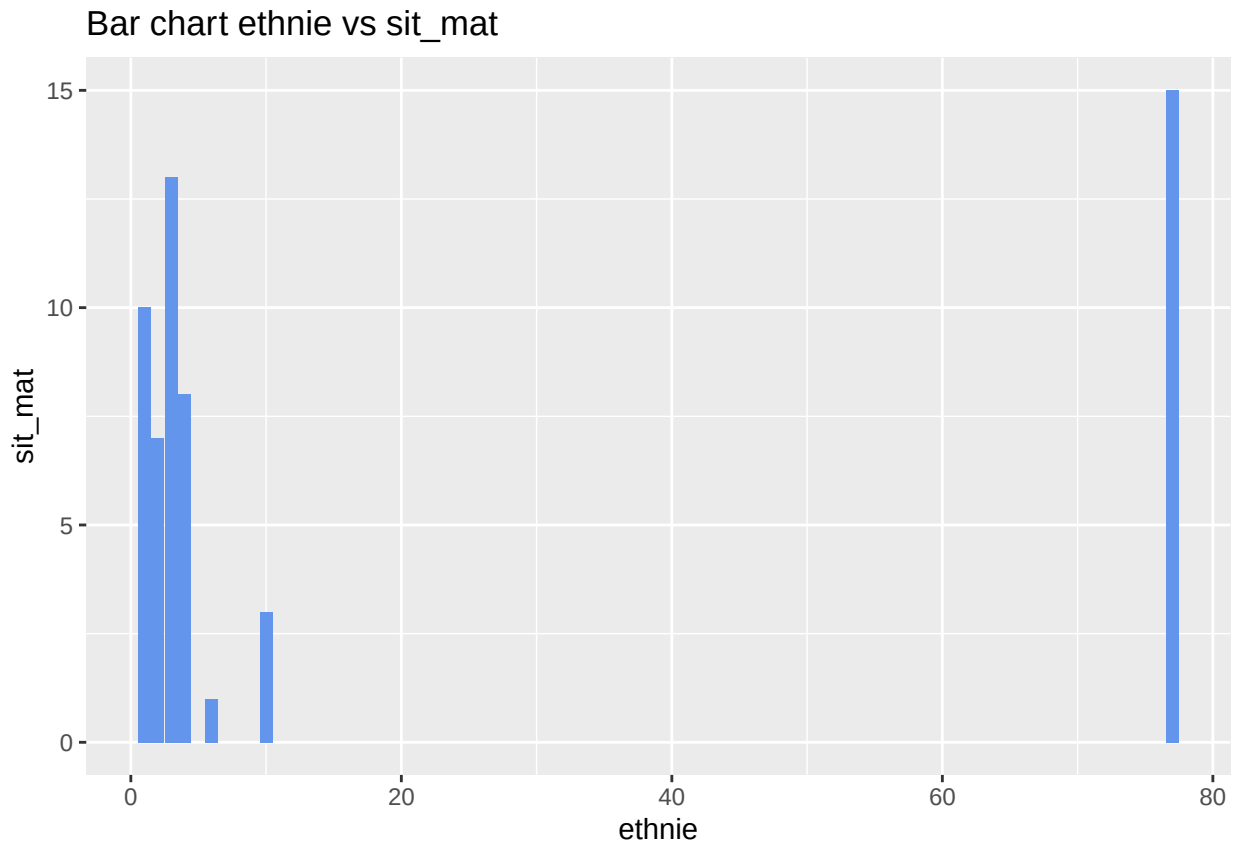
```
##
## $line
```


Line plot sit_mat vs ethnie
ethnie
sit_mat
80
60
40
20
0
1.0
1.5
2.0
2.5
3.0

```
##
## $bar_x
```

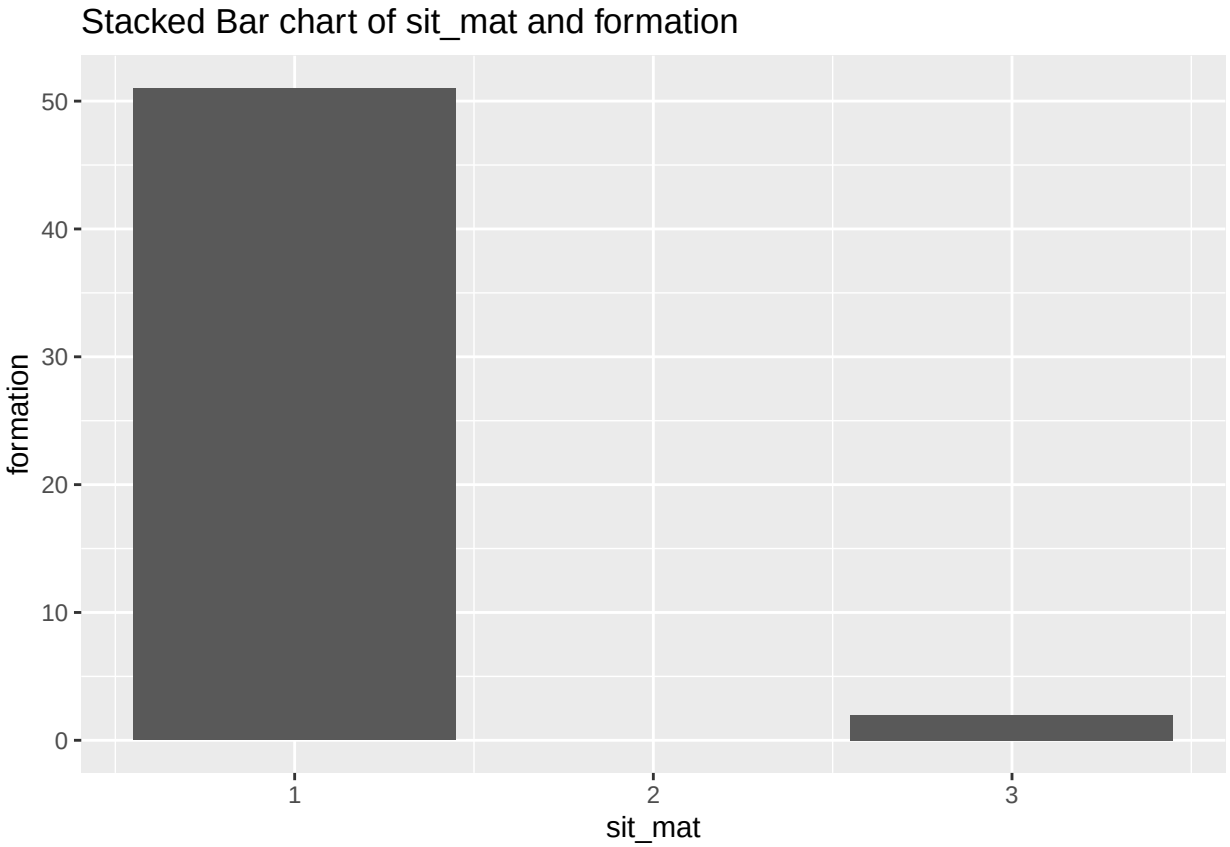

Bar chart sit_mat vs ethnie
ethnie
800
600
400
200
0
0
1
2
3
4
sit_mat

```
##
## $bar_y
```

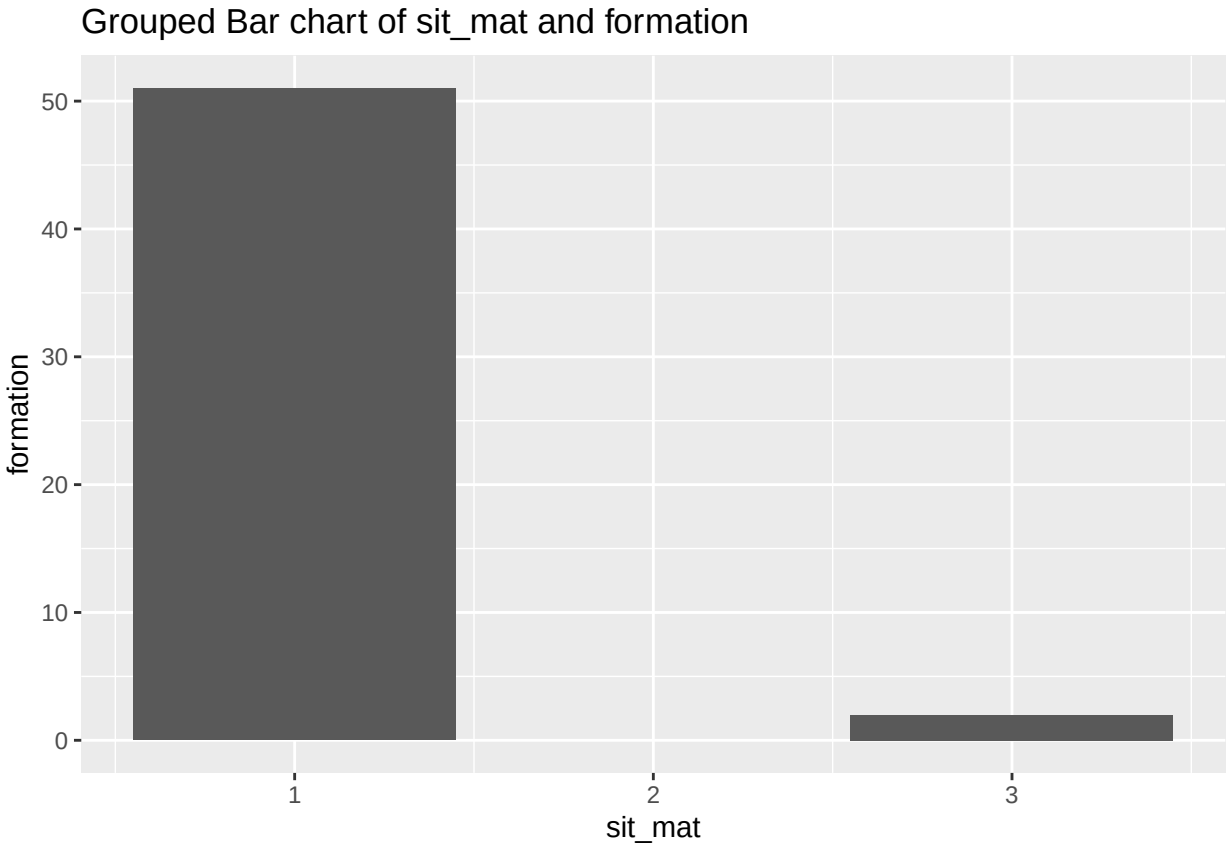



```
bivarie(base_tp2, "sit_mat", "formation", plot = TRUE)
```

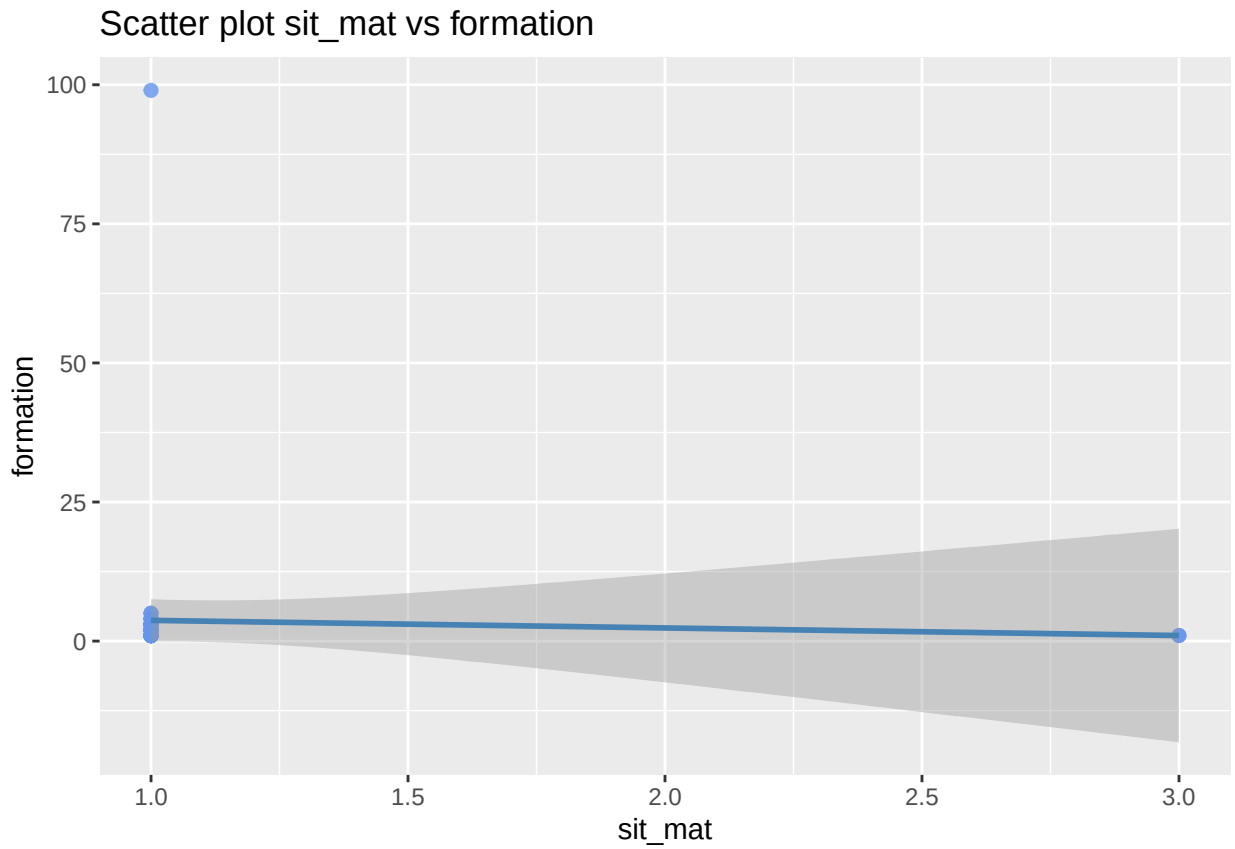
```
##
##      1  2  3  4  5 99
##   1 28 10  7  3  2  1
##   3  2  0  0  0  0  0

## $stacked
```


Stacked Bar chart of sit_mat and formation
formation
50
40
30
20
10
0
1
2
3
sit_mat

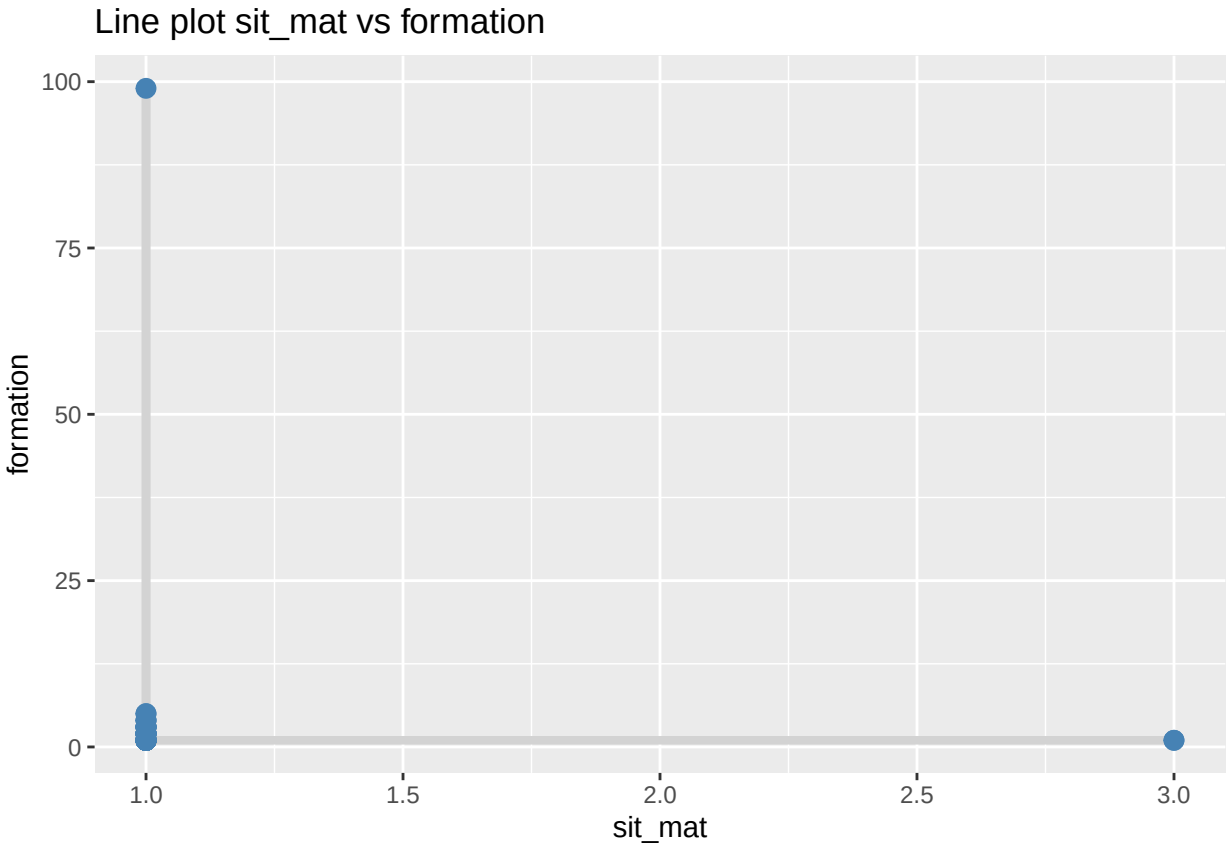
```
##
## $grouped
```


Grouped Bar chart of sit_mat and formation
formation
50
40
30
20
10
0
1
2
3
sit_mat

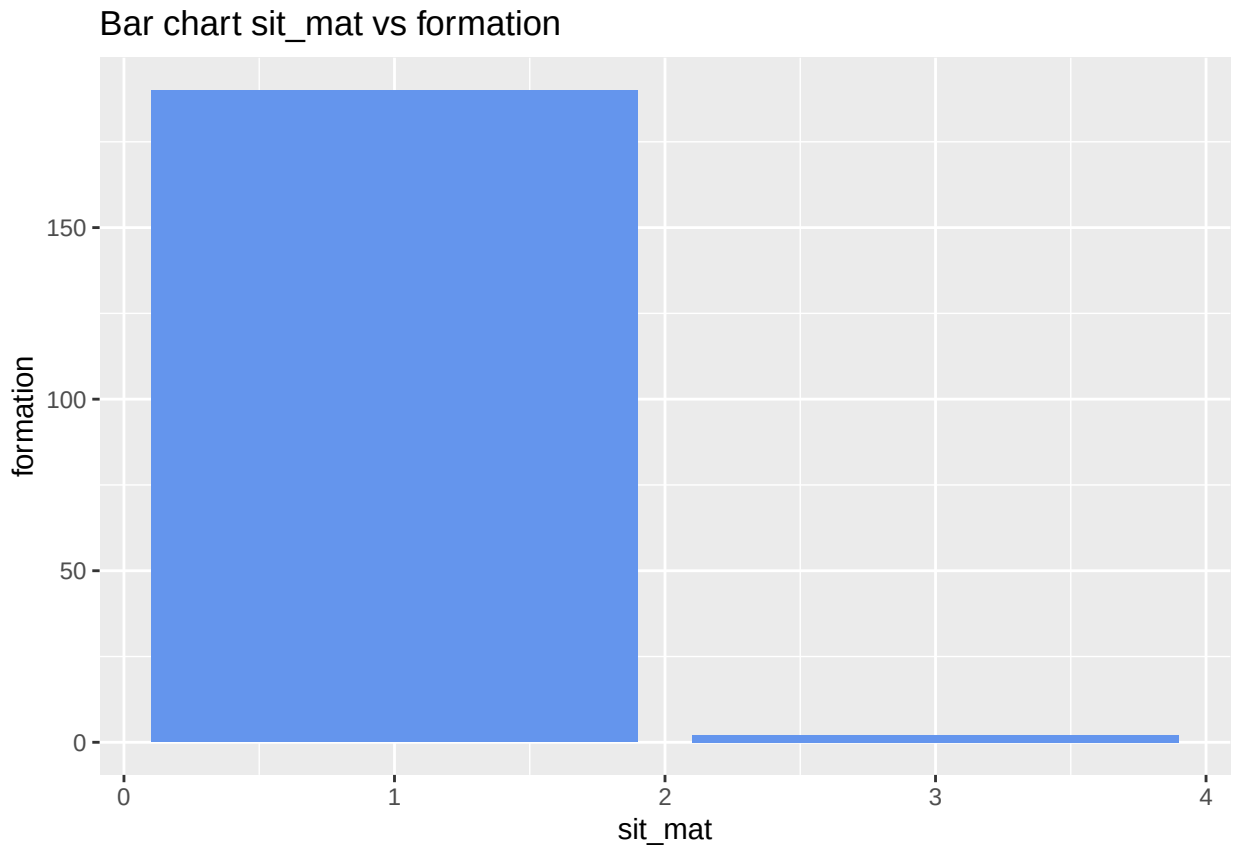
```
##
## $scatter
```


Scatter plot sit_mat vs formation
100
75
50
25
0
formation
1.0
1.5
2.0
2.5
3.0
sit_mat

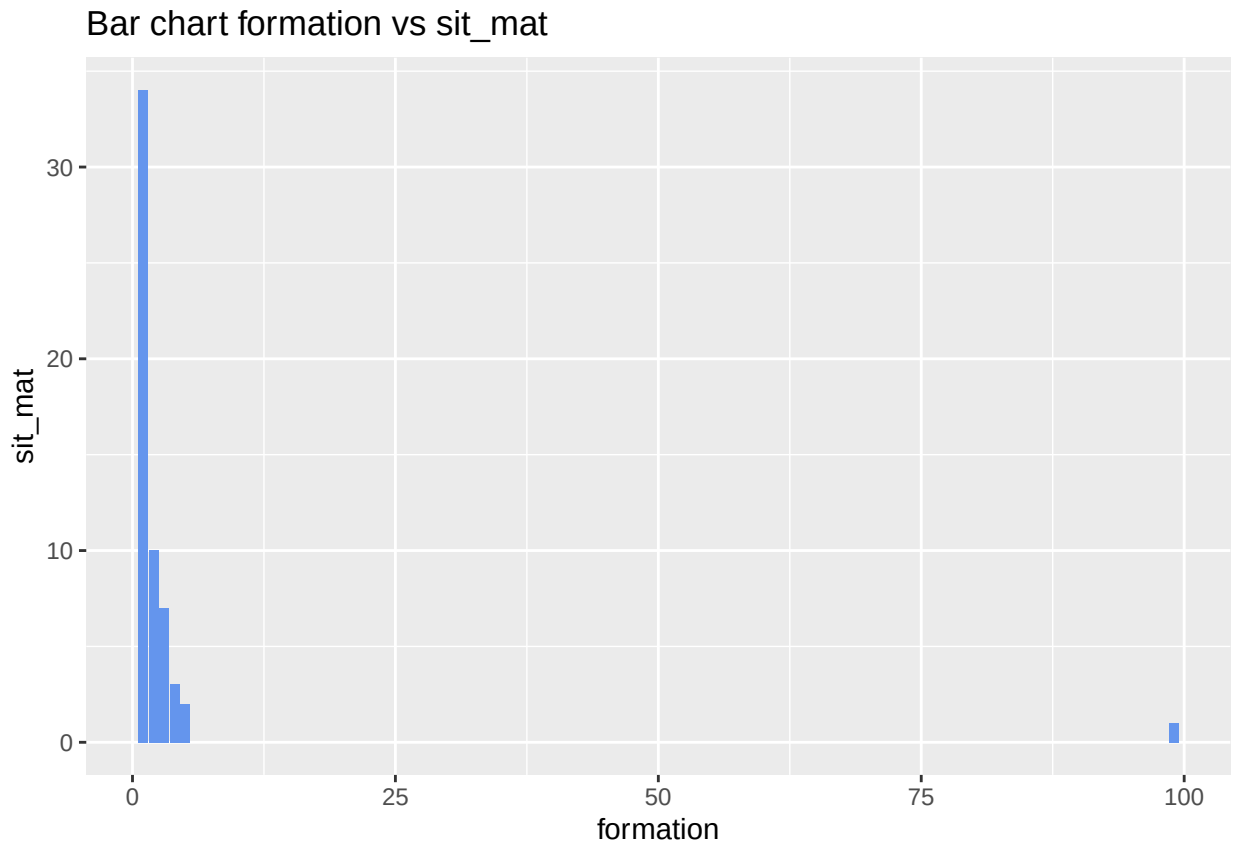
```
##
## $line
```


Line plot sit_mat vs formation
100
75
50
25
0
formation
1.0
1.5
2.0
2.5
3.0
sit_mat

```
##
## $bar_x
```

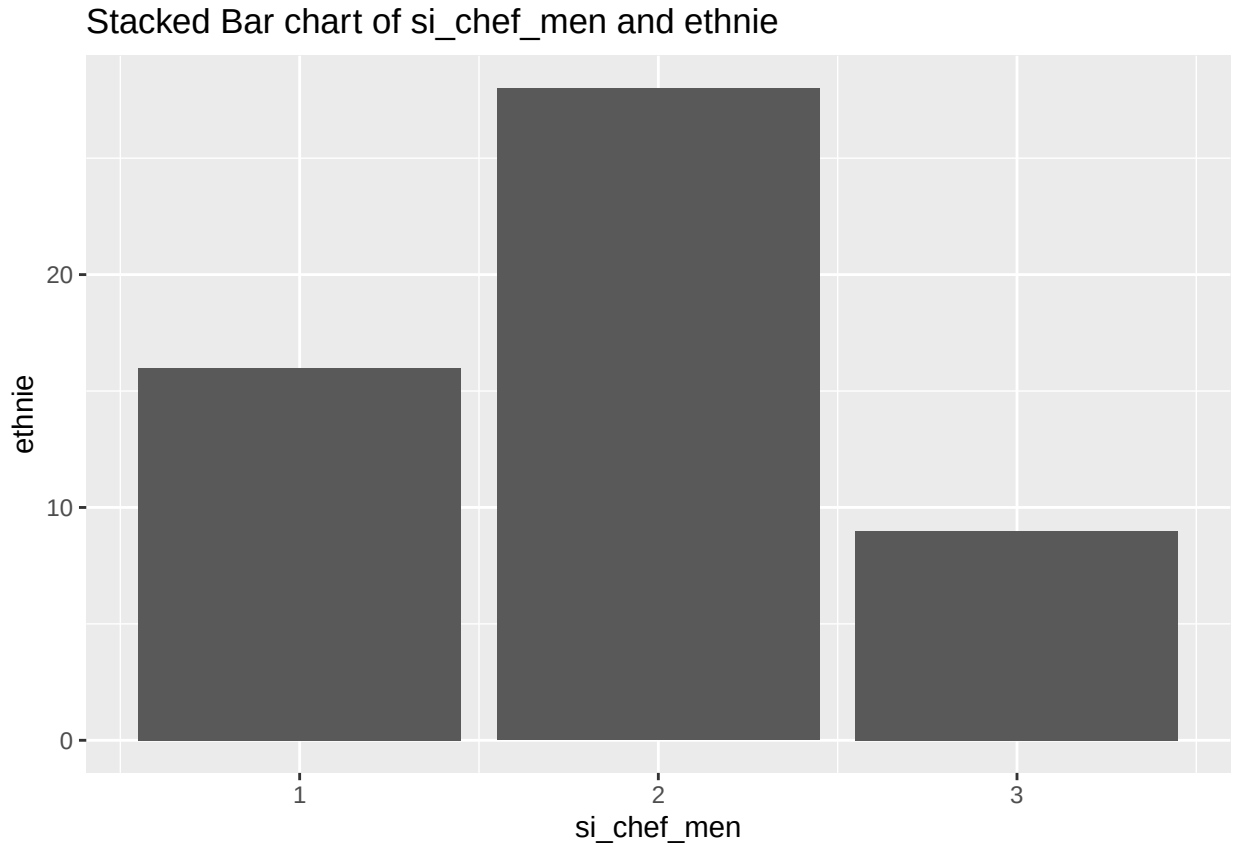

Bar chart sit_mat vs formation
formation
150
100
50
0
0
1
2
3
4
sit_mat

```
##
## $bar_y
```

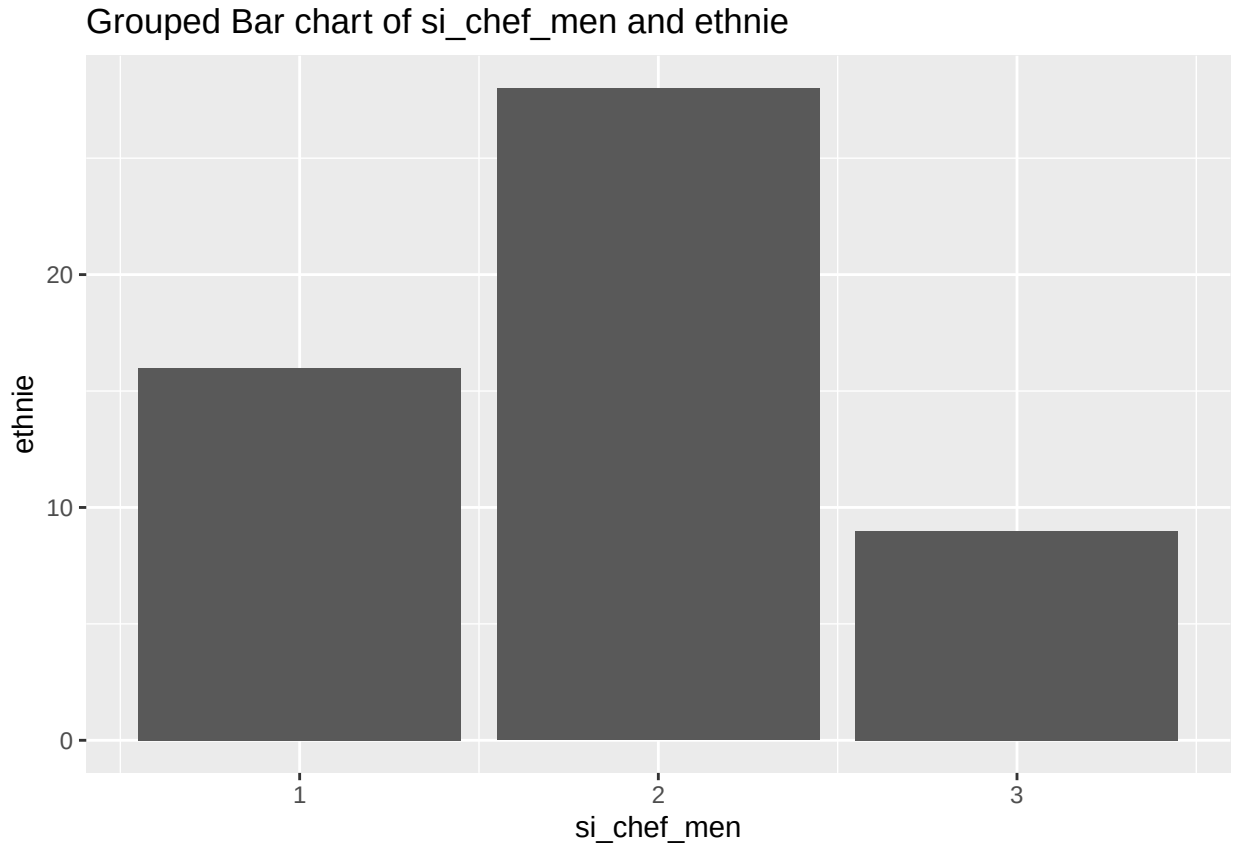



```
bivarie(base_tp2, "si_chef_men", "ethnie", plot = TRUE)
```

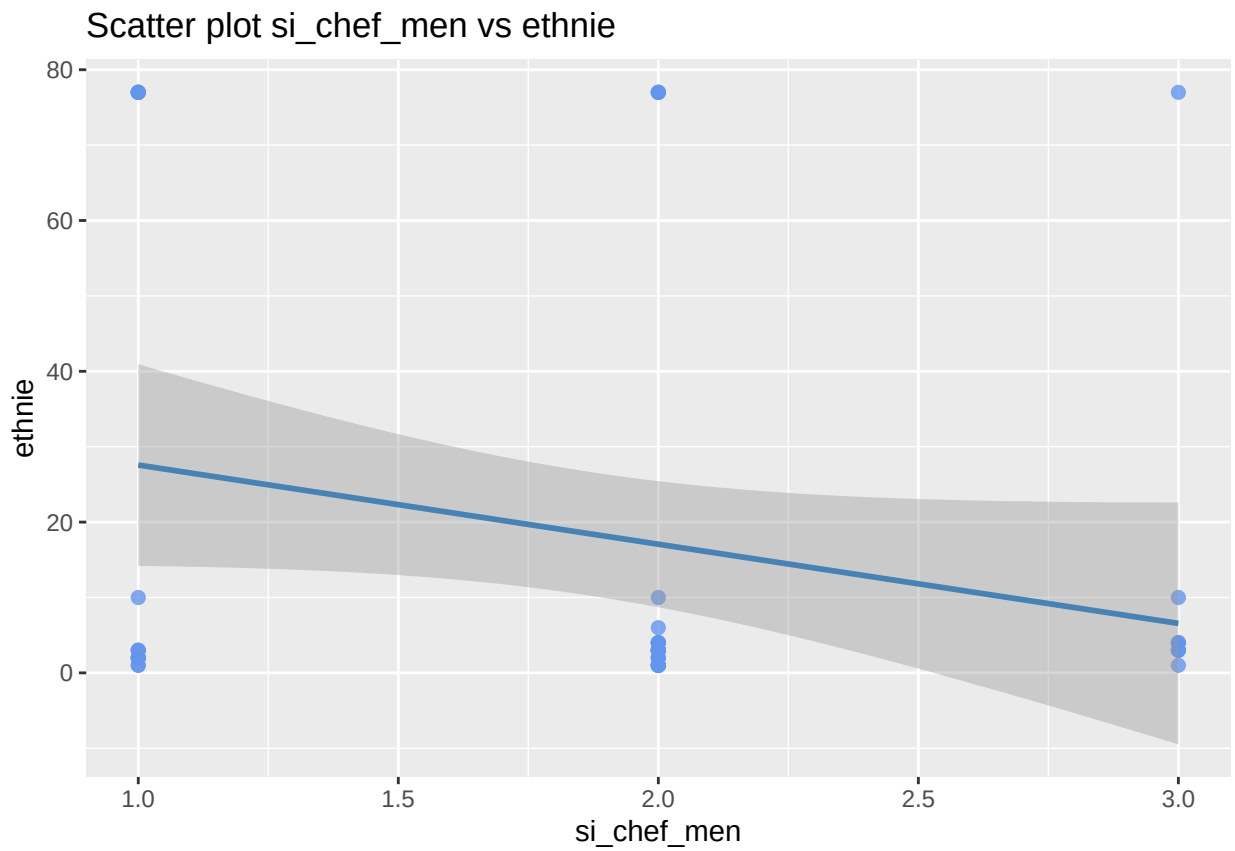
```
##
##     1 2 3 4 6 10 77
##   1 2 4 3 0 0  1  6
##   2 7 3 7 5 1  1  4
##   3 1 0 3 3 0  1  1

## $stacked
```


Stacked Bar chart of si_chef_men and ethnie
ethnie
20
10
0
1
2
3
si_chef_men

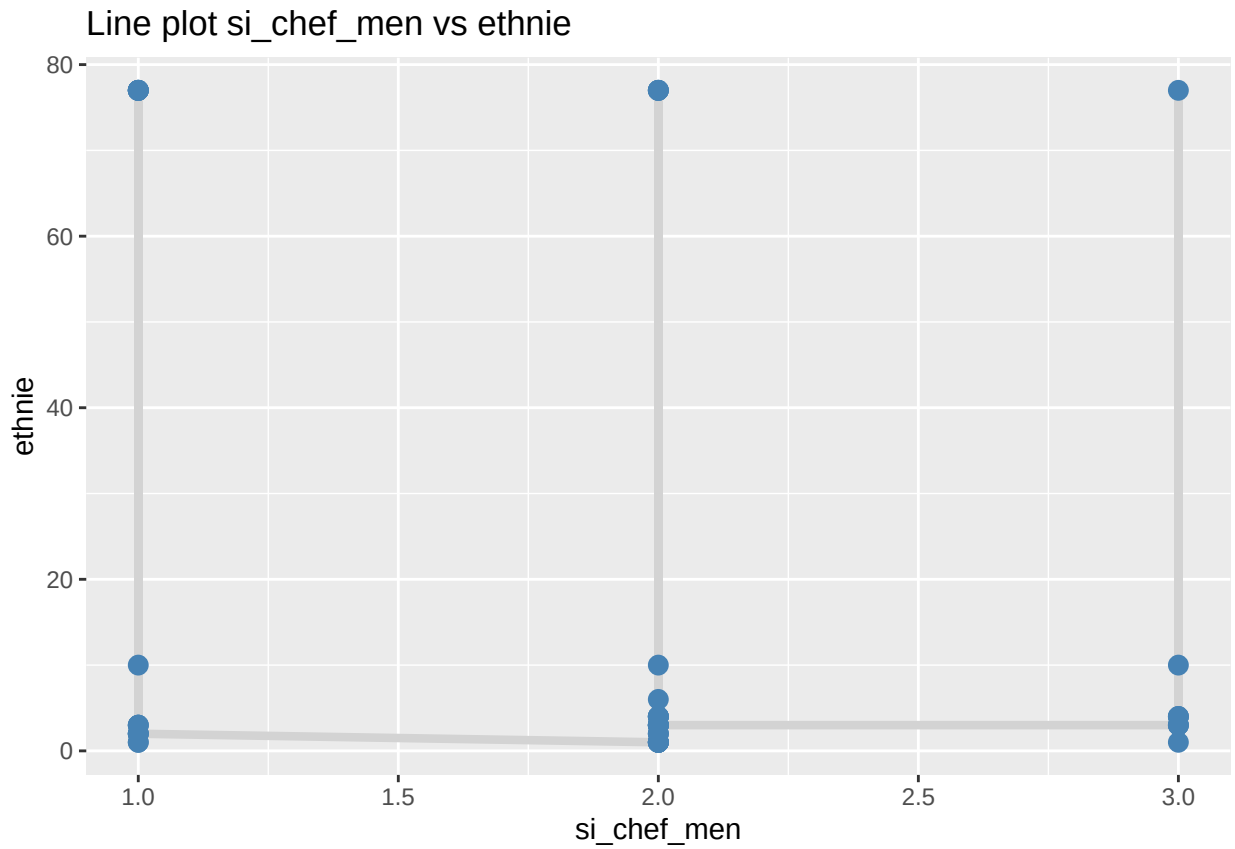
```
##
## $grouped
```


Grouped Bar chart of si_chef_men and ethnie
ethnie
20
10
0
1
2
3
si_chef_men

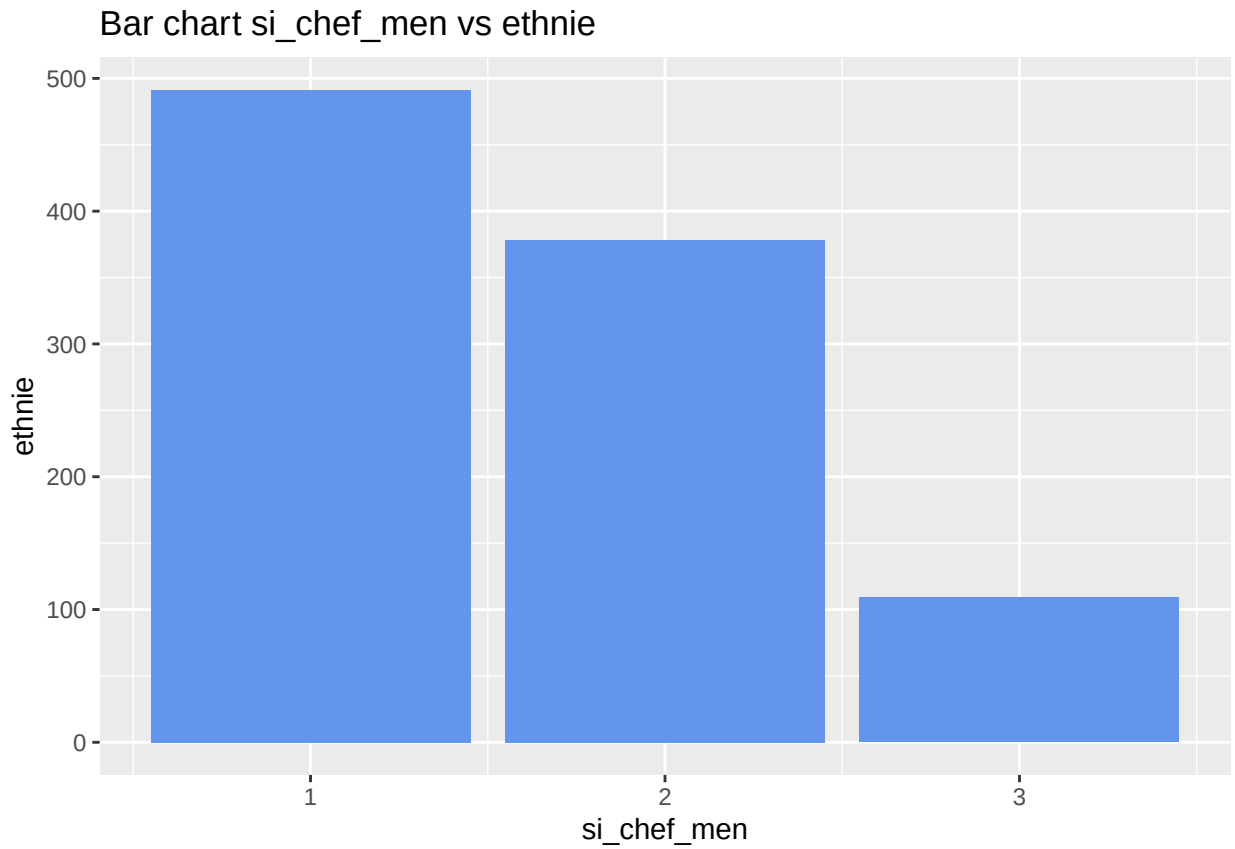
```
##
## $scatter
```


Scatter plot si_chef_men vs ethnie
80
60
40
20
0
ethnie
1.0
1.5
2.0
2.5
3.0
si_chef_men

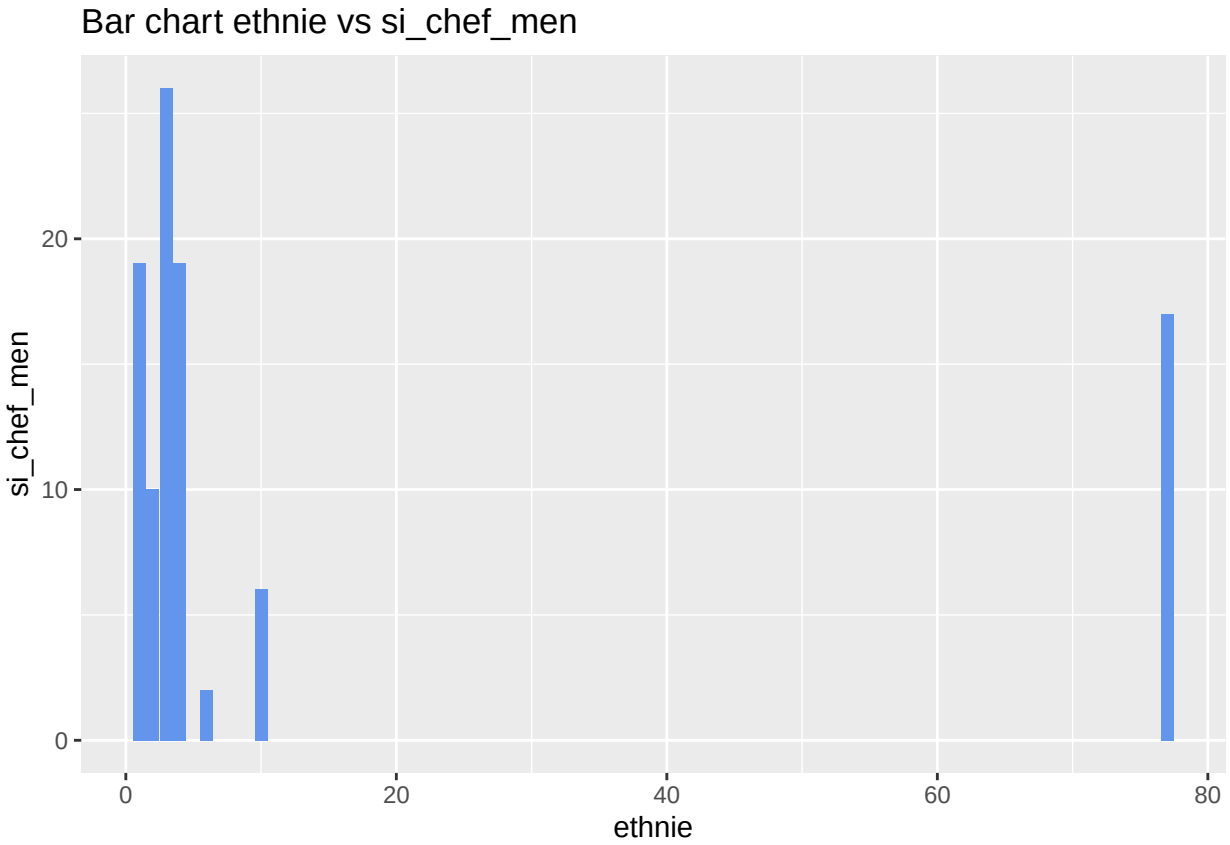
```
##
## $line
```


Line plot si_chef_men vs ethnie
ethnie
80
60
40
20
0
1.0
1.5
2.0
2.5
3.0
si_chef_men

```
##
## $bar_x
```

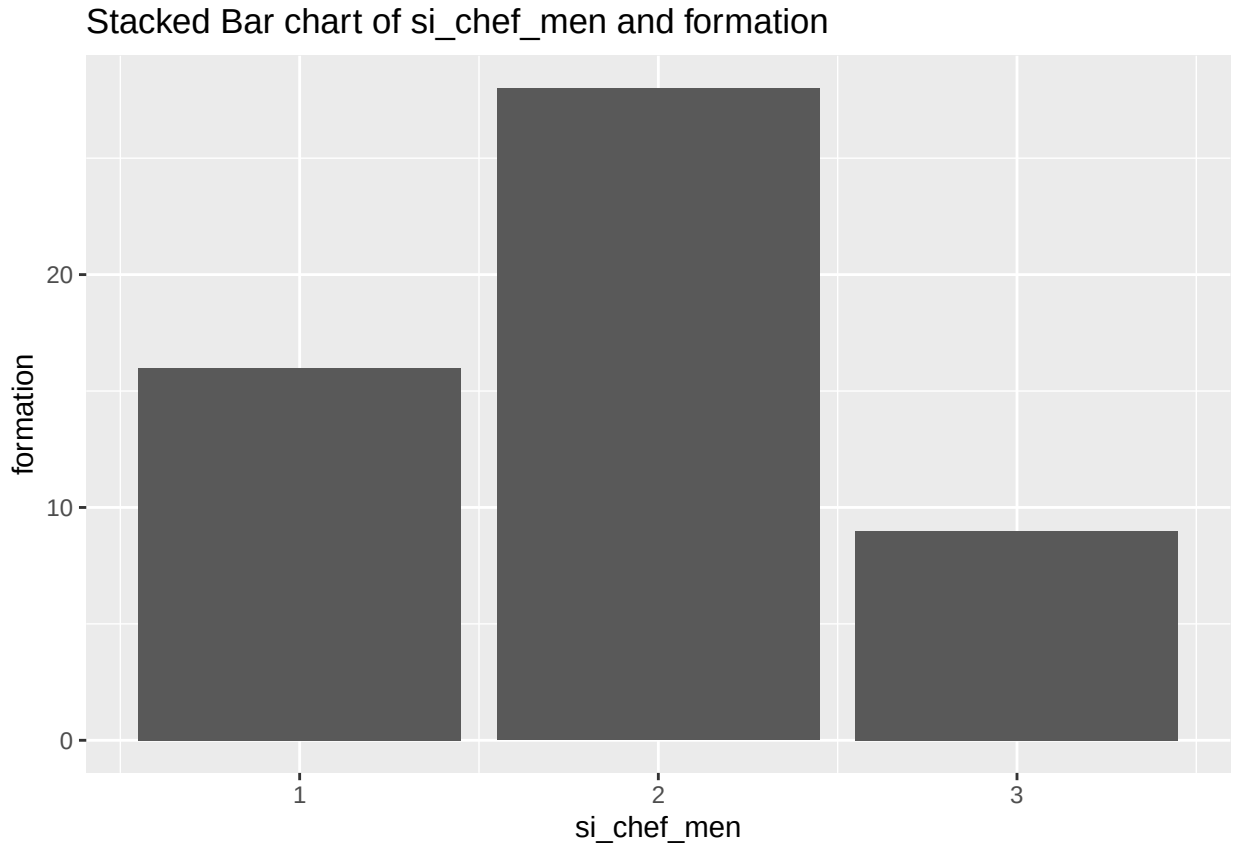

Bar chart si_chef_men vs ethnie
500
400
300
200
100
0
ethnie
1
2
3
si_chef_men

```
##
## $bar_y
```

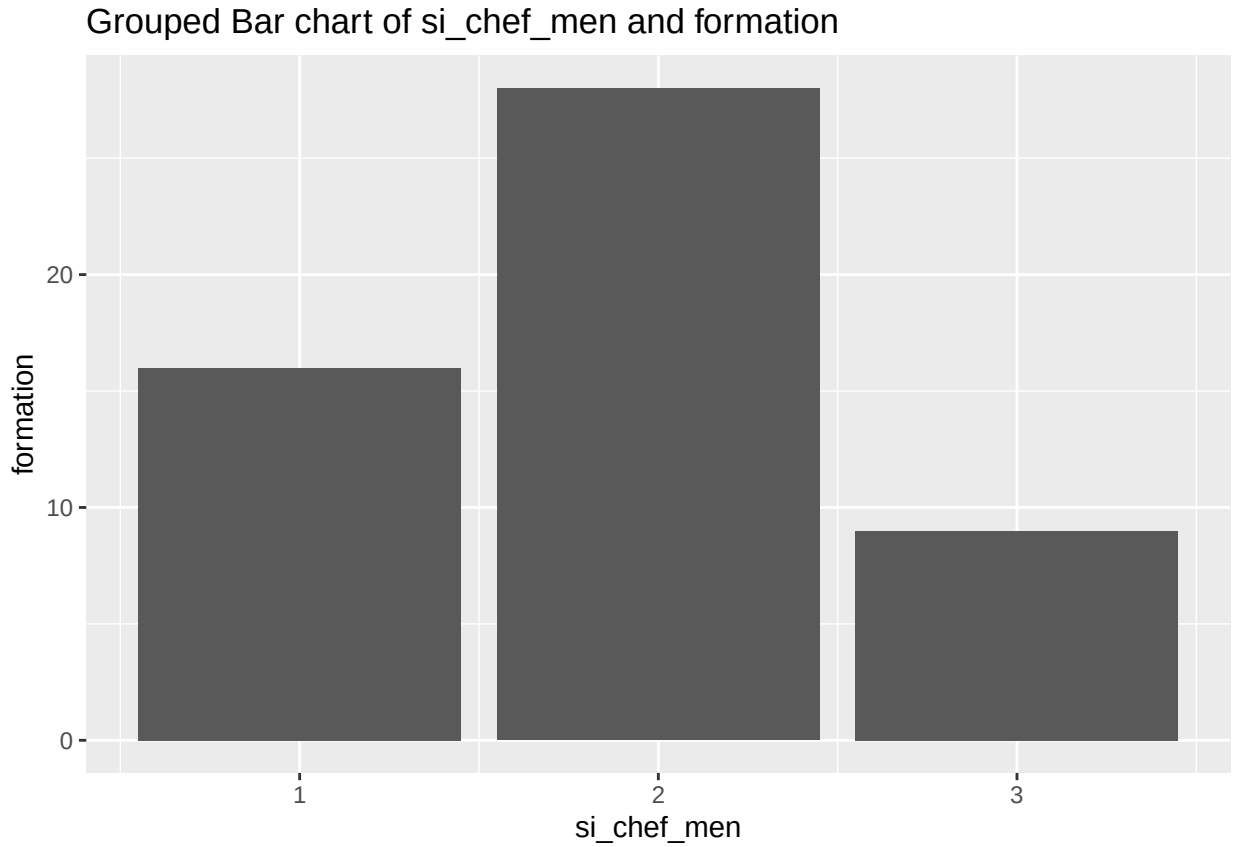



```
bivarie(base_tp2, "si_chef_men", "formation", plot = TRUE)
```

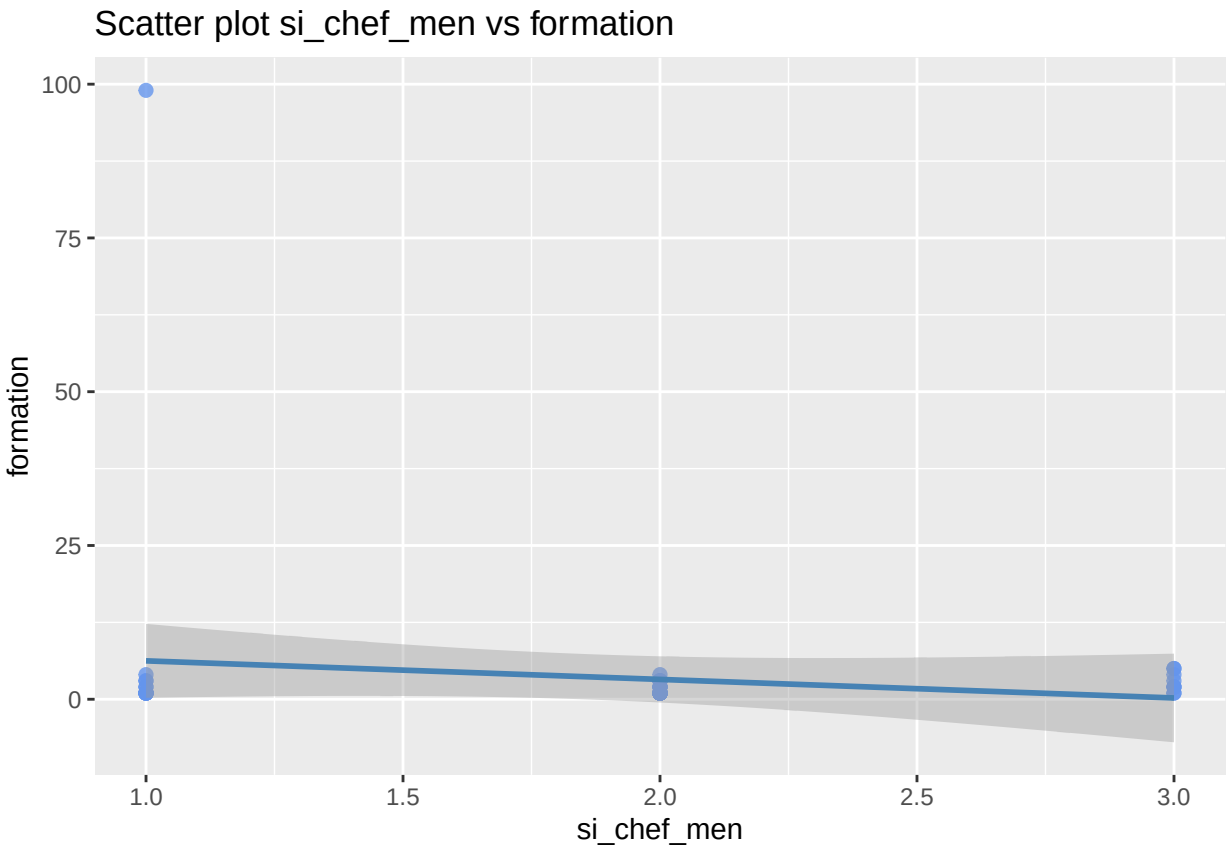
```
##
##      1  2  3  4  5 99
##   1 10  2  2  1  0  1
##   2 18  5  4  1  0  0
##   3  2  3  1  1  2  0

## $stacked
```


Stacked Bar chart of si_chef_men and formation
formation
20
10
0
1
2
3
si_chef_men

```
##
## $grouped
```


Grouped Bar chart of si_chef_men and formation
formation
20
10
0
1
2
3
si_chef_men

```
##
## $scatter
```


Scatter plot si_chef_men vs formation
100
75
50
25
0
formation
1.0
1.5
2.0
2.5
3.0
si_chef_men

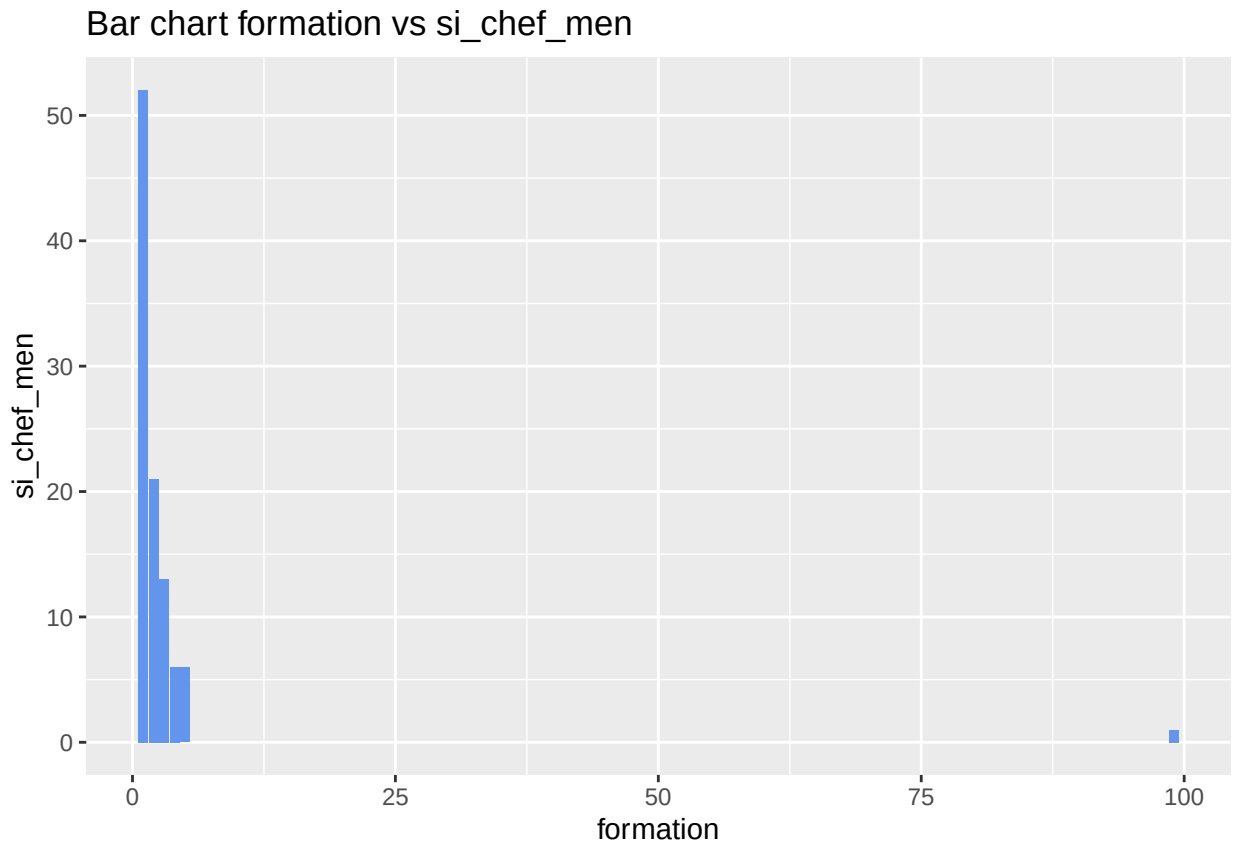
```
##
## $line
```


Line plot si_chef_men vs formation
100
75
50
25
0
formation
1.0
1.5
2.0
2.5
3.0
si_chef_men

```
##
## $bar_x
```

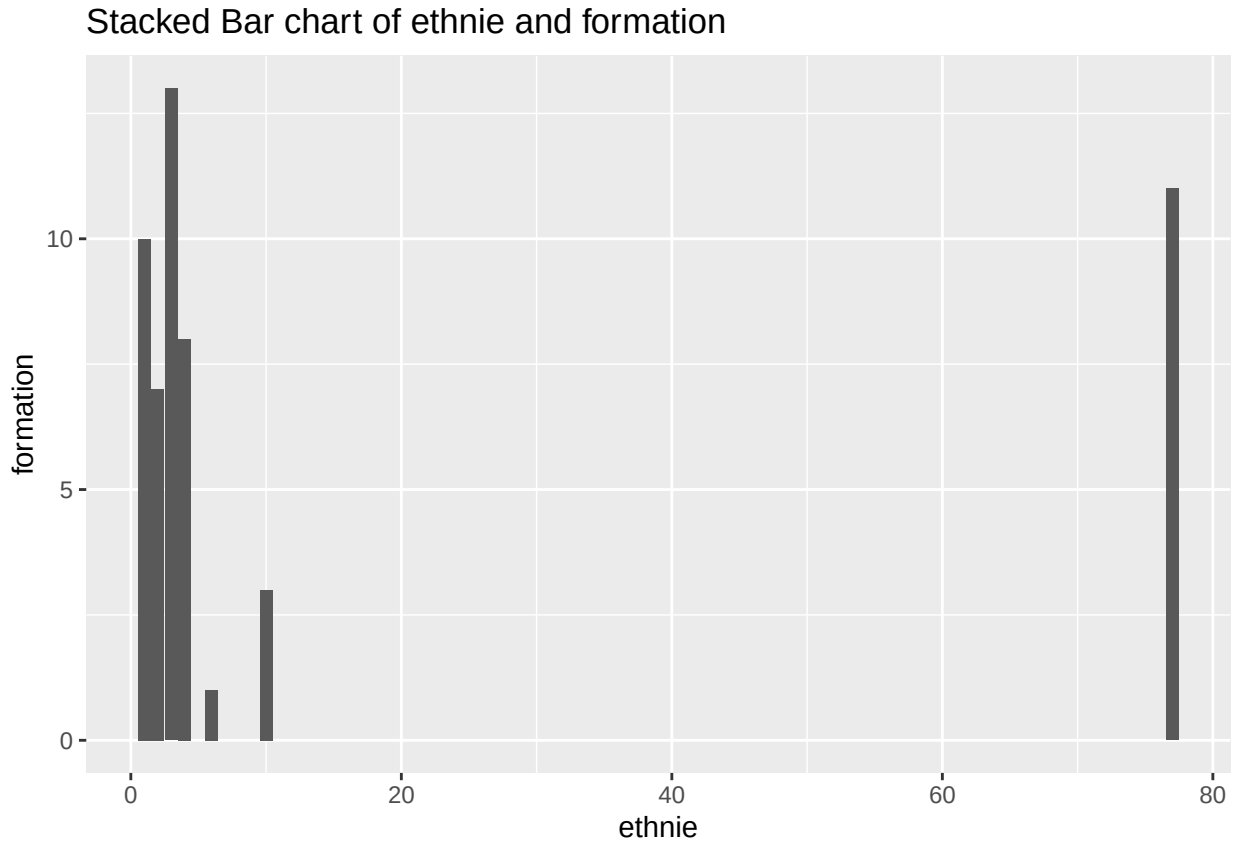

Bar chart si_chef_men vs formation
125
100
75
50
25
0
formation
1
2
3
si_chef_men

```
##
## $bar_y
```

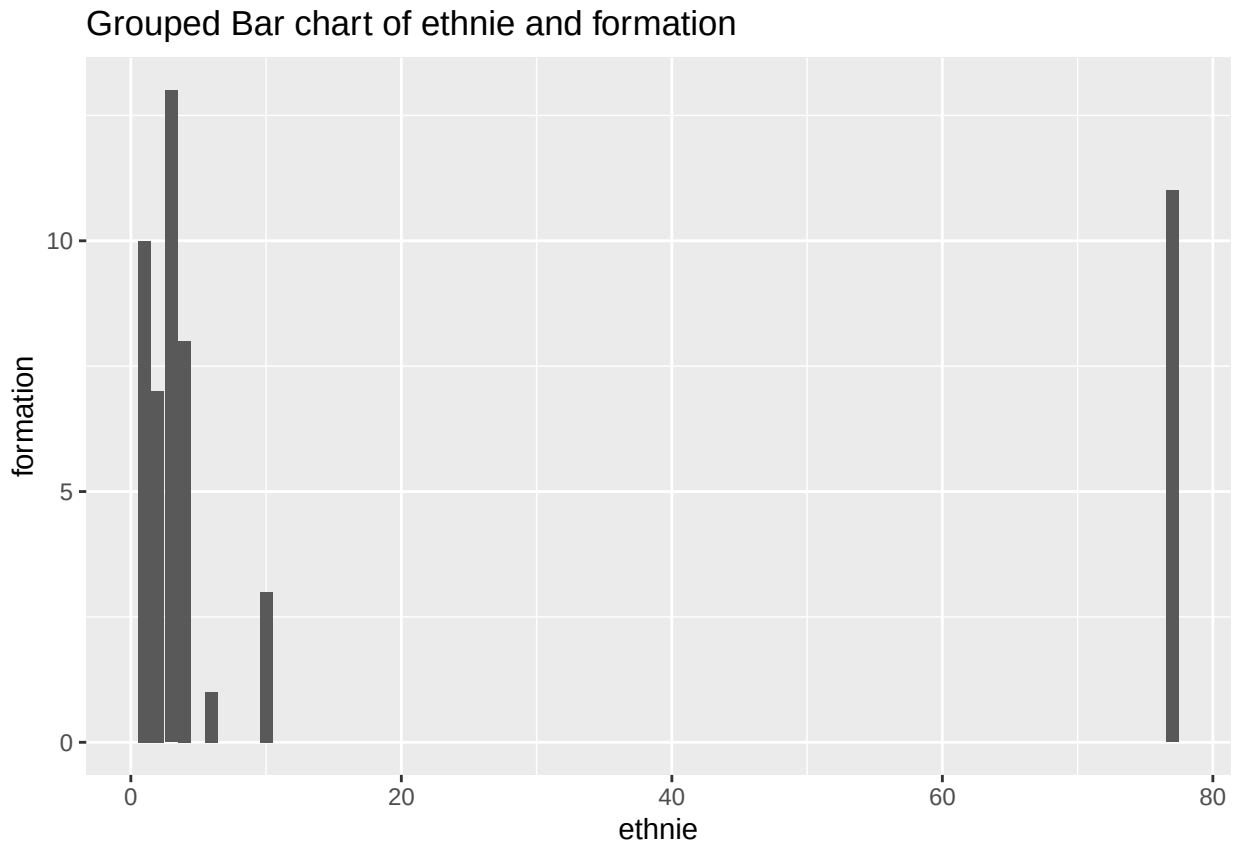



```
bivarie(base_tp2, "ethnie", "formation", plot = TRUE)
```

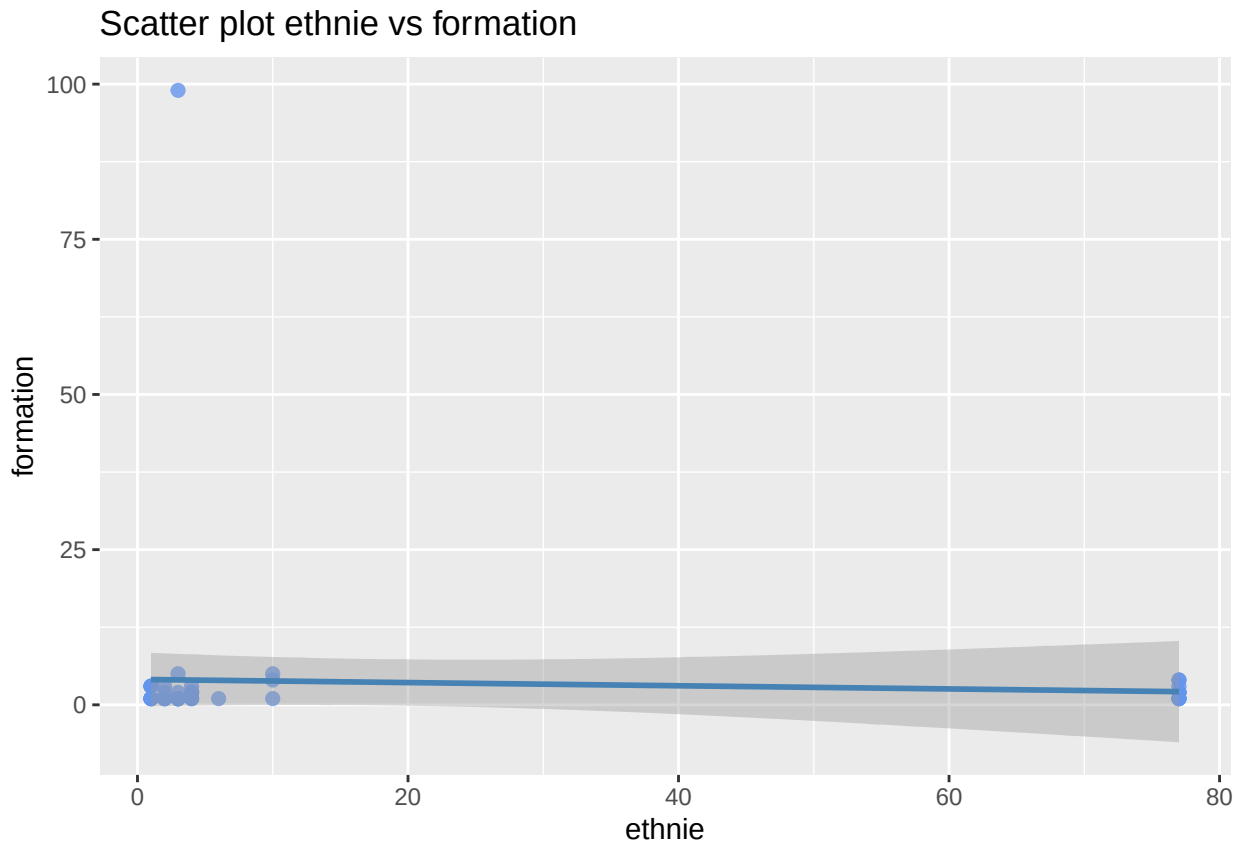
```
##
##       1  2  3  4  5 99
##   1   6  0  4  0  0  0
##   2   5  1  1  0  0  0
##   3  10  1  0  0  1  1
##   4   3  4  1  0  0  0
##   6   1  0  0  0  0  0
##   10  1  0  0  1  1  0
##   77  4  4  1  2  0  0

## $stacked
```


Stacked Bar chart of ethnie and formation
formation
10
5
0
0
20
40
60
80
ethnie

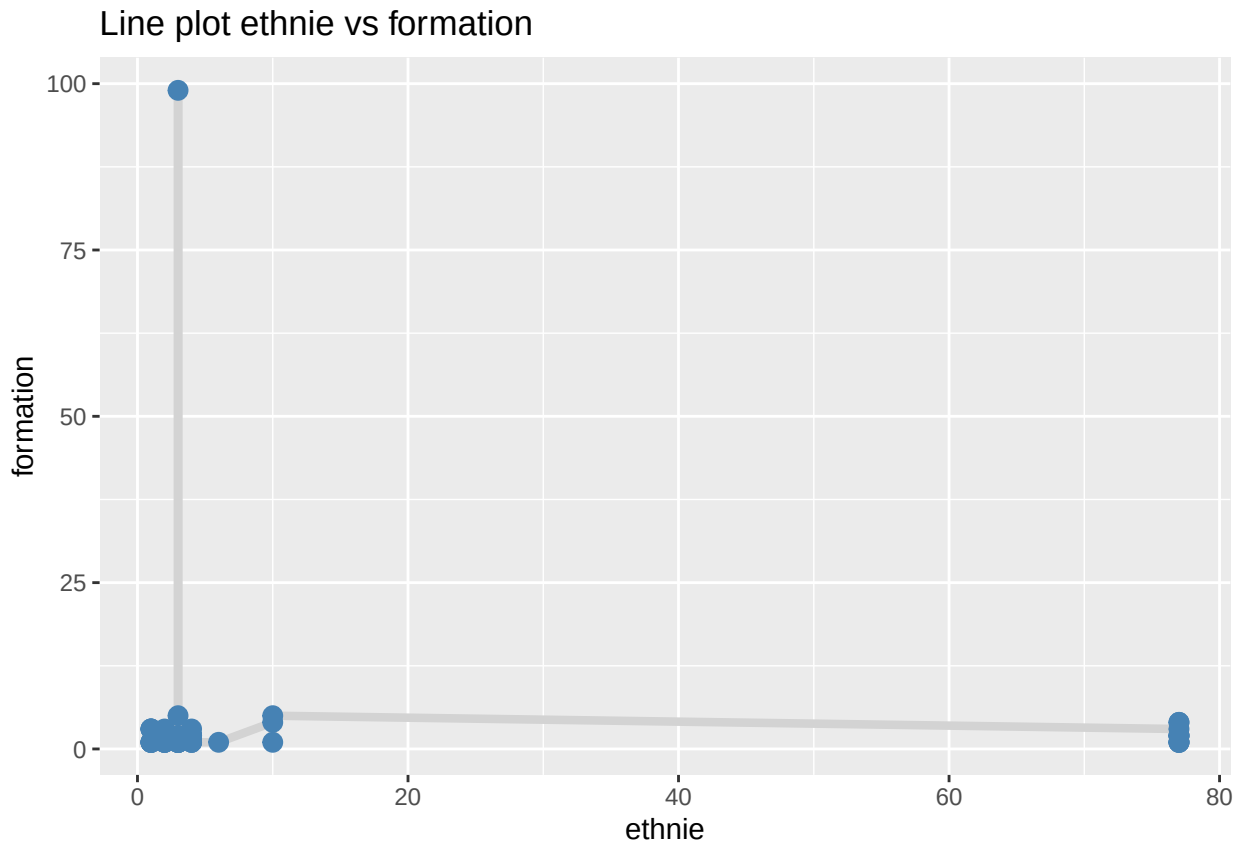
```
##
## $grouped
```


Grouped Bar chart of ethnie and formation
formation
10
5
0
0
20
40
60
80
ethnie

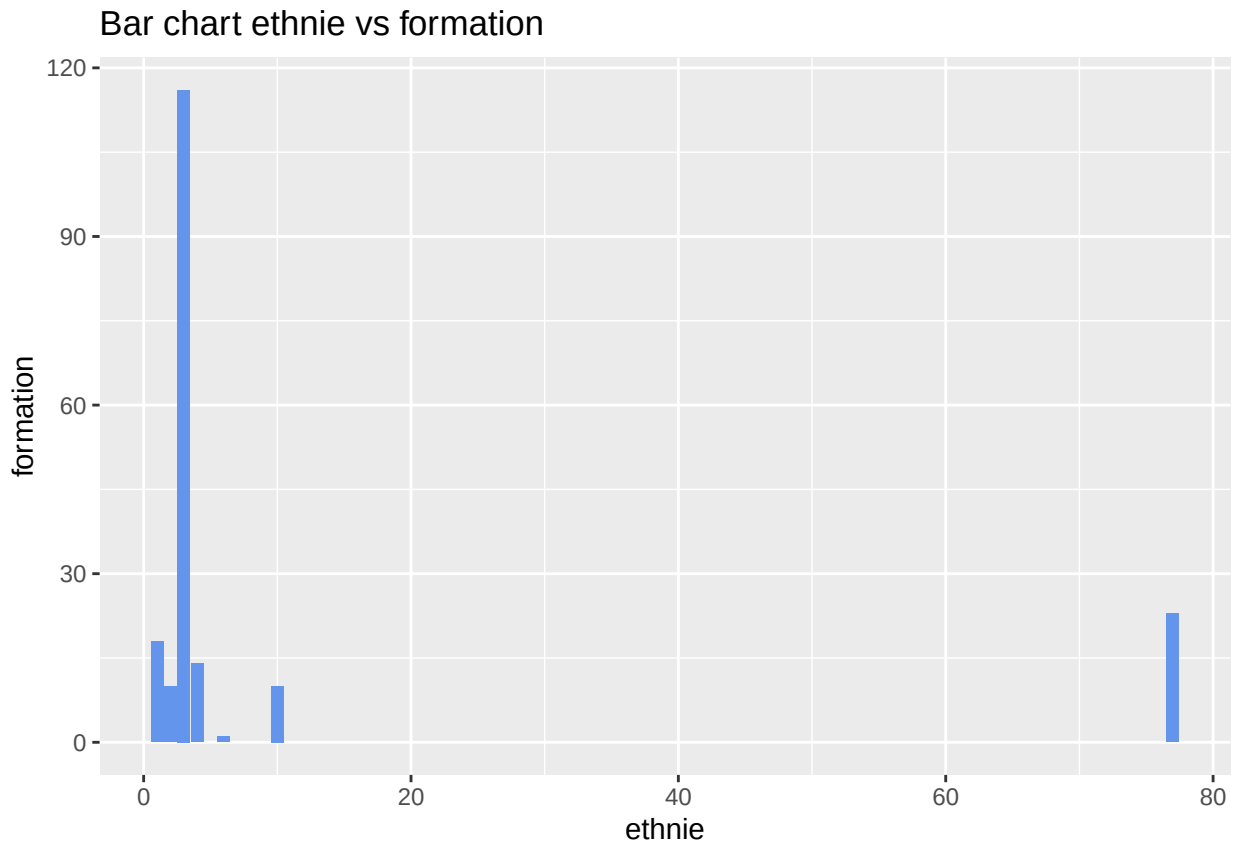
```
##
## $scatter
```


Scatter plot ethnie vs formation
100
75
50
25
0
formation
0
20
40
60
80
ethnie

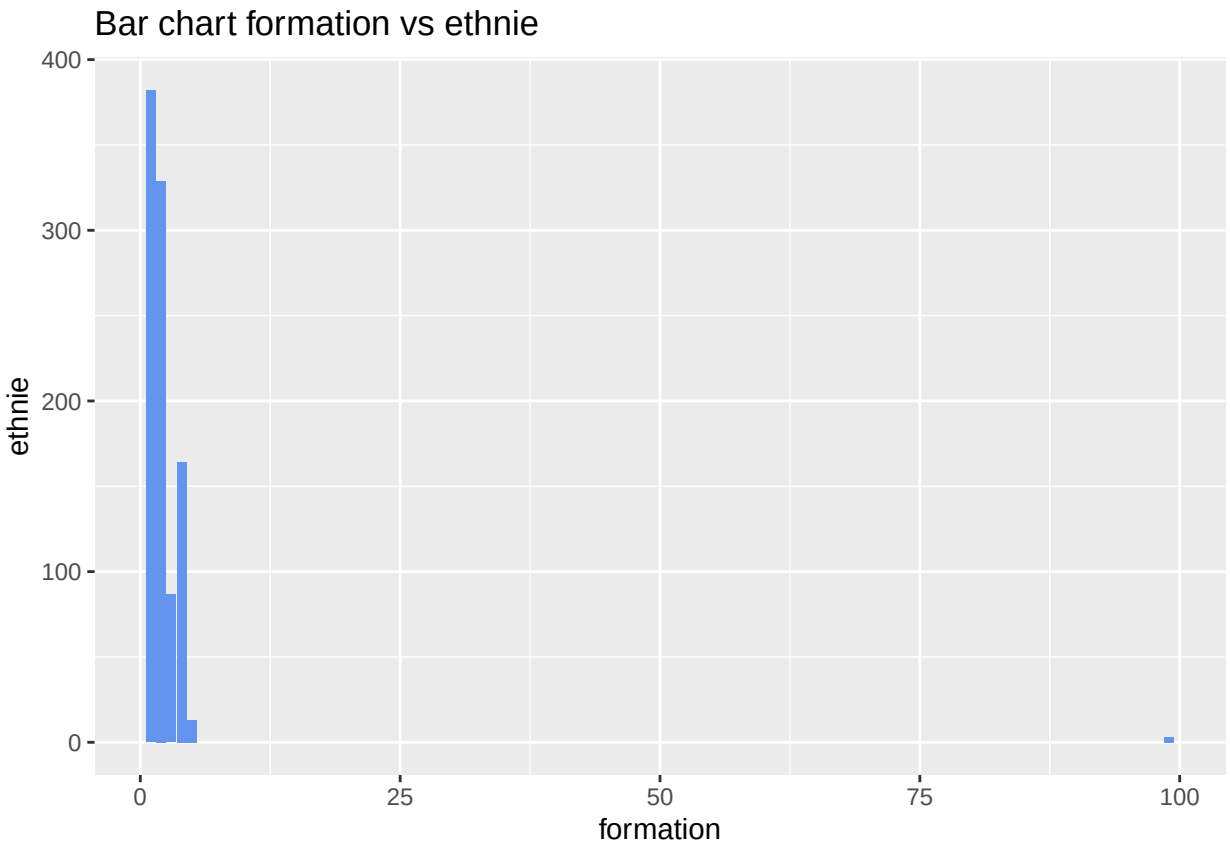
```
##
## $line
```


Line plot ethnie vs formation
100
75
50
25
0
formation
0
20
40
60
80
ethnie

```
##
## $bar_x
```


Bar chart ethnie vs formation
120
90
60
30
0
formation
0
20
40
60
80
ethnie

```
##
## $bar_y
```


Bar chart formation vs ethnie
ethnie
400
300
200
100
0
0
25
50
75
100
formation