Survival analysis of RNAseq for 73 Breast Test

VIVI TANG

17/11/2021

1. Load dataset

```
tucker1=read.csv("/Users/vivit/73TestBreast/RNAseq AE test results for Breast Cancer 73.csv",header=F)
clinical=read.csv("/Users/vivit/73TestBreast/test_clinical_B73_merged.csv",header=T,na.strings=" ")
```

2. Further scale, this step is optional

```
#tucker1=scale(tucker)
```

3. Kmeans 3.1. Apply Kmeans for 2,3 and 4 groups

```
# apply Kmeans
library(flexclust)
```

```
## Warning: package 'lattice' was built under R version 3.6.2
```

```
km2=kcca(tucker1,k=2)
km3=kcca(tucker1,k=3)
km4=kcca(tucker1,k=4)
# add kmeans results to the datasets
data1=clinical
data1$km2=km2@cluster
data1$km3=km3@cluster
data1$km4=km4@cluster
```

3.2. Create the survival objects

```
# Create survival object
library(survival)
```

```
## Warning: package 'survival' was built under R version 3.6.2
```

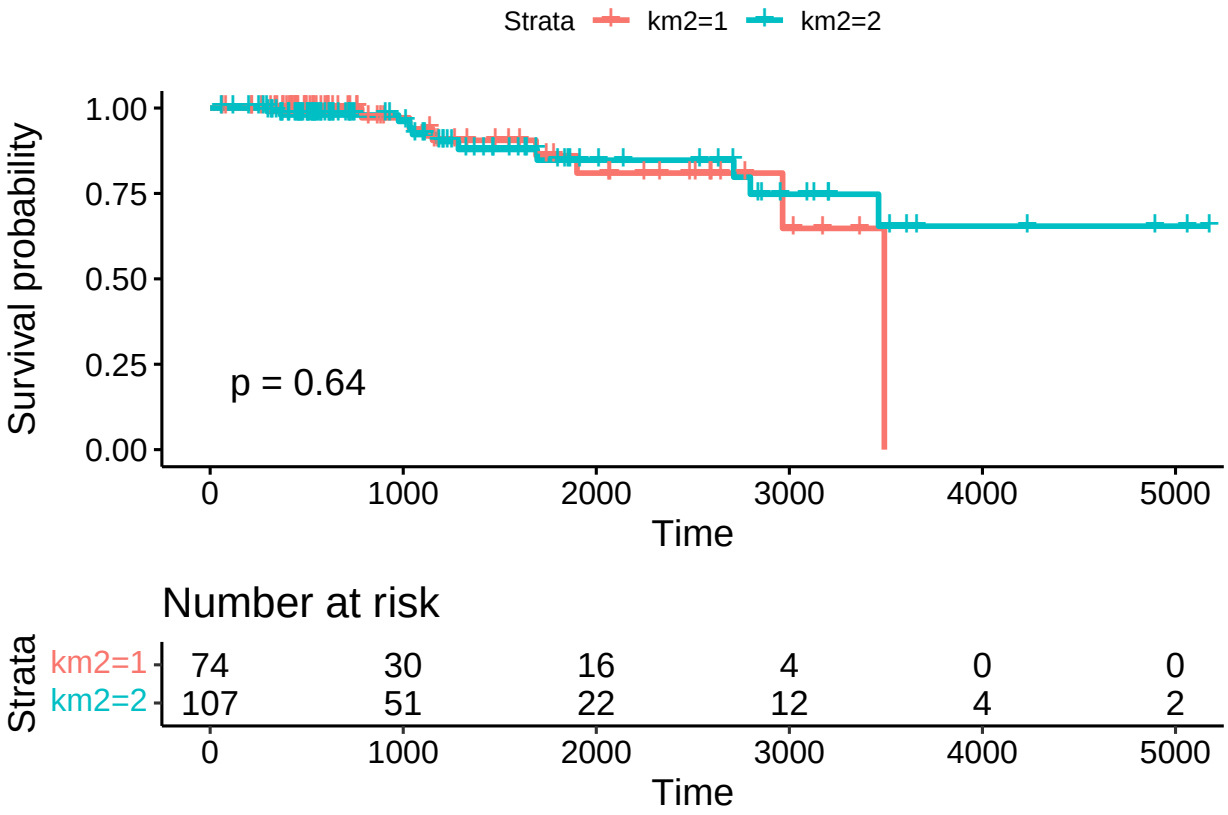
```
library(survminer)
```

```
## Warning: package 'survminer' was built under R version 3.6.2

## Warning: package 'ggplot2' was built under R version 3.6.2

## Warning: package 'ggpubr' was built under R version 3.6.2
```

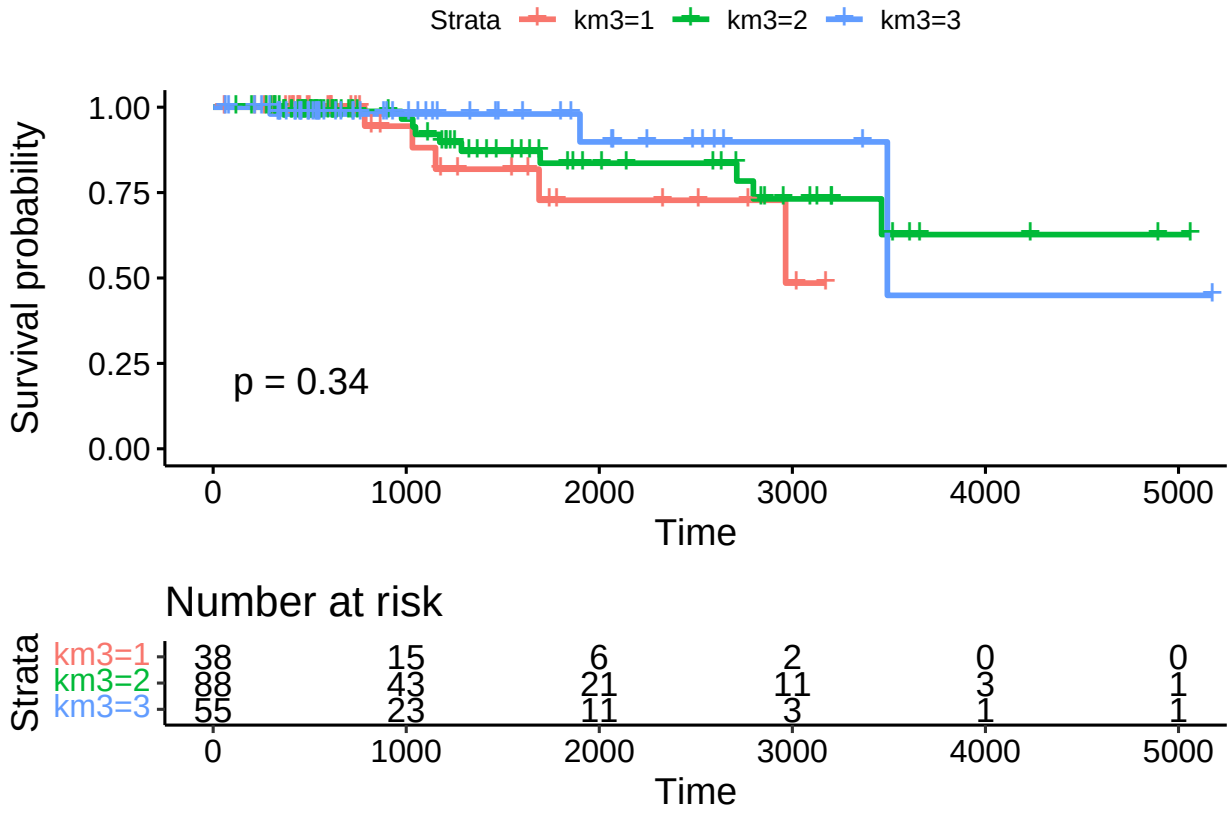
```
obj=Surv(time=data1$overall_survival,event=data1$status)
```

3.3. Fit the results of Kmeans

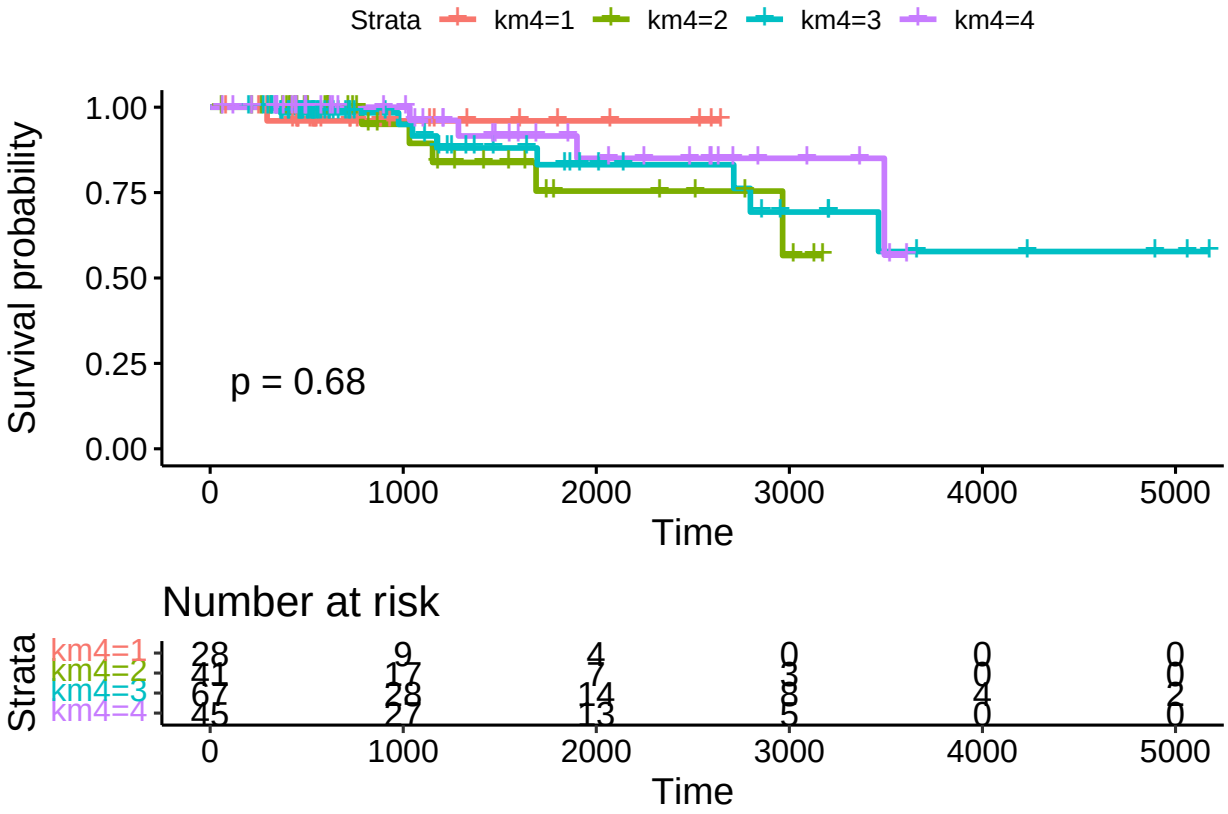
```
km_2 = survfit(obj ~ km2, data = data1)
ggsurvplot(km_2,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
km_3= survfit(obj ~ km3, data = data1)
ggsurvplot(km_3,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
km_4 = survfit(obj ~ km4, data = data1)
ggsurvplot(km_4,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```



4. Hierarchical Clustering 4.1.Apply all different combinations of distance measurement and linkage methods for Hierarchical clustering

```
#Apply all different combinations of distance measurement and linkage methods for Hierarchical clusteri
library(proxy)
```

```
## Warning: package 'proxy' was built under R version 3.6.2
```

```
ds1=dist(tucker1,"euclidean")
ks11=hclust(ds1,"single")
ks12=hclust(ds1,"average")
ks13=hclust(ds1,"complete")
ks14=hclust(ds1,"ward")
ks15=hclust(ds1,"mcquitty")
ks16=hclust(ds1,"median")
ks17=hclust(ds1,"centroid")

ds2=dist(tucker1,"manhattan")
ks21=hclust(ds2,"single")
ks22=hclust(ds2,"average")
ks23=hclust(ds2,"complete")
ks24=hclust(ds2,"ward")
ks25=hclust(ds2,"mcquitty")
ks26=hclust(ds2,"median")
ks27=hclust(ds2,"centroid")

ds3=dist(tucker1,"Jaccard")
```

```
ks31=hclust(ds3,"single")
ks32=hclust(ds3,"average")
ks33=hclust(ds3,"complete")
ks34=hclust(ds3,"ward")
ks35=hclust(ds3,"mcquitty")
ks36=hclust(ds3,"median")
ks37=hclust(ds3,"centroid")

ds4=dist(tucker1,"maximum")
ks41=hclust(ds4,"single")
ks42=hclust(ds4,"average")
ks43=hclust(ds4,"complete")
ks44=hclust(ds4,"ward")
ks45=hclust(ds4,"mcquitty")
ks46=hclust(ds4,"median")
ks47=hclust(ds4,"centroid")

ds5=dist(tucker1,"binary")
ks51=hclust(ds5,"single")
ks52=hclust(ds5,"average")
ks53=hclust(ds5,"complete")
ks54=hclust(ds5,"ward")
ks55=hclust(ds5,"mcquitty")
ks56=hclust(ds5,"median")
ks57=hclust(ds5,"centroid")

ds6=dist(tucker1,"canberra")
ks61=hclust(ds6,"single")
ks62=hclust(ds6,"average")
ks63=hclust(ds6,"complete")
ks64=hclust(ds6,"ward")
ks65=hclust(ds6,"mcquitty")
ks66=hclust(ds6,"median")
ks67=hclust(ds6,"centroid")

euc_s2=cutree(ks11,k=2)
euc_c2=cutree(ks12,k=2)
euc_a2=cutree(ks13,k=2)
euc_w2=cutree(ks14,k=2)
euc_mcq2=cutree(ks15,k=2)
euc_med2=cutree(ks16,k=2)
euc_cen2=cutree(ks17,k=2)

man_s2=cutree(ks21,k=2)
man_c2=cutree(ks22,k=2)
man_a2=cutree(ks23,k=2)
man_w2=cutree(ks24,k=2)
man_mcq2=cutree(ks25,k=2)
man_med2=cutree(ks26,k=2)
man_cen2=cutree(ks27,k=2)

Jacc_s2=cutree(ks31,k=2)
Jacc_c2=cutree(ks32,k=2)
```

```
Jacc_a2=cutree(ks33,k=2)
Jacc_w2=cutree(ks34,k=2)
Jacc_mcq2=cutree(ks35,k=2)
Jacc_med2=cutree(ks36,k=2)
Jacc_cen2=cutree(ks37,k=2)

max_s2=cutree(ks41,k=2)
max_c2=cutree(ks42,k=2)
max_a2=cutree(ks43,k=2)
max_w2=cutree(ks44,k=2)
max_mcq2=cutree(ks45,k=2)
max_med2=cutree(ks46,k=2)
max_cen2=cutree(ks47,k=2)

bin_s2=cutree(ks51,k=2)
bin_c2=cutree(ks52,k=2)
bin_a2=cutree(ks53,k=2)
bin_w2=cutree(ks54,k=2)
bin_mcq2=cutree(ks55,k=2)
bin_med2=cutree(ks56,k=2)
bin_cen2=cutree(ks57,k=2)

can_s2=cutree(ks61,k=2)
can_c2=cutree(ks62,k=2)
can_a2=cutree(ks63,k=2)
can_w2=cutree(ks64,k=2)
can_mcq2=cutree(ks65,k=2)
can_med2=cutree(ks66,k=2)
can_cen2=cutree(ks67,k=2)

# add all clustered results into the dataset
data1$euc_s2=euc_s2
data1$euc_c2=euc_c2
data1$euc_a2=euc_a2
data1$euc_w2=euc_w2
data1$euc_mcq2=euc_mcq2
data1$euc_med2=euc_med2
data1$euc_cen2=euc_cen2

data1$man_s2=man_s2
data1$man_c2=man_c2
data1$man_a2=man_a2
data1$man_w2=man_w2
data1$man_mcq2=man_mcq2
data1$man_med2=man_med2
data1$man_cen2=man_cen2

data1$Jacc_s2=Jacc_s2
data1$Jacc_c2=Jacc_c2
data1$Jacc_a2=Jacc_a2
data1$Jacc_w2=Jacc_w2
data1$Jacc_mcq2=Jacc_mcq2
data1$Jacc_med2=Jacc_med2
```

```
data1$Jacc_cen2=Jacc_cen2

data1$max_s2=max_s2
data1$max_c2=max_c2
data1$max_a2=max_a2
data1$max_w2=max_w2
data1$max_mcq2=max_mcq2
data1$max_med2=max_med2
data1$max_cen2=max_cen2

data1$bin_s2=bin_s2
data1$bin_c2=bin_c2
data1$bin_a2=bin_a2
data1$bin_w2=bin_w2
data1$bin_mcq2=bin_mcq2
data1$bin_med2=bin_med2
data1$bin_cen2=bin_cen2

data1$can_s2=can_s2
data1$can_c2=can_c2
data1$can_a2=can_a2
data1$can_w2=can_w2
data1$can_mcq2=can_mcq2
data1$can_med2=can_med2
data1$can_cen2=can_cen2


euc_s3=cutree(ks11,k=3)
euc_c3=cutree(ks12,k=3)
euc_a3=cutree(ks13,k=3)
euc_w3=cutree(ks14,k=3)
euc_mcq3=cutree(ks15,k=3)
euc_med3=cutree(ks16,k=3)
euc_cen3=cutree(ks17,k=3)

man_s3=cutree(ks21,k=3)
man_c3=cutree(ks22,k=3)
man_a3=cutree(ks23,k=3)
man_w3=cutree(ks24,k=3)
man_mcq3=cutree(ks25,k=3)
man_med3=cutree(ks26,k=3)
man_cen3=cutree(ks27,k=3)

Jacc_s3=cutree(ks31,k=3)
Jacc_c3=cutree(ks32,k=3)
Jacc_a3=cutree(ks33,k=3)
Jacc_w3=cutree(ks34,k=3)
Jacc_mcq3=cutree(ks35,k=3)
Jacc_med3=cutree(ks36,k=3)
Jacc_cen3=cutree(ks37,k=3)

max_s3=cutree(ks41,k=3)
max_c3=cutree(ks42,k=3)
```

```
max_a3=cutree(ks43,k=3)
max_w3=cutree(ks44,k=3)
max_mcq3=cutree(ks45,k=3)
max_med3=cutree(ks46,k=3)
max_cen3=cutree(ks47,k=3)

bin_s3=cutree(ks51,k=3)
bin_c3=cutree(ks52,k=3)
bin_a3=cutree(ks53,k=3)
bin_w3=cutree(ks54,k=3)
bin_mcq3=cutree(ks55,k=3)
bin_med3=cutree(ks56,k=3)
bin_cen3=cutree(ks57,k=3)

can_s3=cutree(ks61,k=3)
can_c3=cutree(ks62,k=3)
can_a3=cutree(ks63,k=3)
can_w3=cutree(ks64,k=3)
can_mcq3=cutree(ks65,k=3)
can_med3=cutree(ks66,k=3)
can_cen3=cutree(ks67,k=3)

data1$euc_s3=euc_s3
data1$euc_c3=euc_c3
data1$euc_a3=euc_a3
data1$euc_w3=euc_w3
data1$euc_mcq3=euc_mcq3
data1$euc_med3=euc_med3
data1$euc_cen3=euc_cen3

data1$man_s3=man_s3
data1$man_c3=man_c3
data1$man_a3=man_a3
data1$man_w3=man_w3
data1$man_mcq3=man_mcq3
data1$man_med3=man_med3
data1$man_cen3=man_cen3

data1$Jacc_s3=Jacc_s3
data1$Jacc_c3=Jacc_c3
data1$Jacc_a3=Jacc_a3
data1$Jacc_w3=Jacc_w3
data1$Jacc_mcq3=Jacc_mcq3
data1$Jacc_med3=Jacc_med3
data1$Jacc_cen3=Jacc_cen3

data1$max_s3=max_s3
data1$max_c3=max_c3
data1$max_a3=max_a3
data1$max_w3=max_w3
data1$max_mcq3=max_mcq3
data1$max_med3=max_med3
data1$max_cen3=max_cen3
```

```
data1$bin_s3=bin_s3
data1$bin_c3=bin_c3
data1$bin_a3=bin_a3
data1$bin_w3=bin_w3
data1$bin_mcq3=bin_mcq3
data1$bin_med3=bin_med3
data1$bin_cen3=bin_cen3

data1$can_s3=can_s3
data1$can_c3=can_c3
data1$can_a3=can_a3
data1$can_w3=can_w3
data1$can_mcq3=can_mcq3
data1$can_med3=can_med3
data1$can_cen3=can_cen3



euc_s4=cutree(ks11,k=4)
euc_c4=cutree(ks12,k=4)
euc_a4=cutree(ks13,k=4)
euc_w4=cutree(ks14,k=4)
euc_mcq4=cutree(ks15,k=4)
euc_med4=cutree(ks16,k=4)
euc_cen4=cutree(ks17,k=4)

man_s4=cutree(ks21,k=4)
man_c4=cutree(ks22,k=4)
man_a4=cutree(ks23,k=4)
man_w4=cutree(ks24,k=4)
man_mcq4=cutree(ks25,k=4)
man_med4=cutree(ks26,k=4)
man_cen4=cutree(ks27,k=4)

Jacc_s4=cutree(ks31,k=4)
Jacc_c4=cutree(ks32,k=4)
Jacc_a4=cutree(ks33,k=4)
Jacc_w4=cutree(ks34,k=4)
Jacc_mcq4=cutree(ks35,k=4)
Jacc_med4=cutree(ks36,k=4)
Jacc_cen4=cutree(ks37,k=4)

max_s4=cutree(ks41,k=4)
max_c4=cutree(ks42,k=4)
max_a4=cutree(ks43,k=4)
max_w4=cutree(ks44,k=4)
max_mcq4=cutree(ks45,k=4)
max_med4=cutree(ks46,k=4)
max_cen4=cutree(ks47,k=4)

bin_s4=cutree(ks51,k=4)
bin_c4=cutree(ks52,k=4)
bin_a4=cutree(ks53,k=4)
```

```
bin_w4=cutree(ks54,k=4)
bin_mcq4=cutree(ks55,k=4)
bin_med4=cutree(ks56,k=4)
bin_cen4=cutree(ks57,k=4)

can_s4=cutree(ks61,k=4)
can_c4=cutree(ks62,k=4)
can_a4=cutree(ks63,k=4)
can_w4=cutree(ks64,k=4)
can_mcq4=cutree(ks65,k=4)
can_med4=cutree(ks66,k=4)
can_cen4=cutree(ks67,k=4)

data1$euc_s4=euc_s4
data1$euc_c4=euc_c4
data1$euc_a4=euc_a4
data1$euc_w4=euc_w4
data1$euc_mcq4=euc_mcq4
data1$euc_med4=euc_med4
data1$euc_cen4=euc_cen4

data1$man_s4=man_s4
data1$man_c4=man_c4
data1$man_a4=man_a4
data1$man_w4=man_w4
data1$man_mcq4=man_mcq4
data1$man_med4=man_med4
data1$man_cen4=man_cen4

data1$Jacc_s4=Jacc_s4
data1$Jacc_c4=Jacc_c4
data1$Jacc_a4=Jacc_a4
data1$Jacc_w4=Jacc_w4
data1$Jacc_mcq4=Jacc_mcq4
data1$Jacc_med4=Jacc_med4
data1$Jacc_cen4=Jacc_cen4

data1$max_s4=max_s4
data1$max_c4=max_c4
data1$max_a4=max_a4
data1$max_w4=max_w4
data1$max_mcq4=max_mcq4
data1$max_med4=max_med4
data1$max_cen4=max_cen4

data1$bin_s4=bin_s4
data1$bin_c4=bin_c4
data1$bin_a4=bin_a4
data1$bin_w4=bin_w4
data1$bin_mcq4=bin_mcq4
data1$bin_med4=bin_med4
data1$bin_cen4=bin_cen4
```

```
data1$can_s4=can_s4
data1$can_c4=can_c4
data1$can_a4=can_a4
data1$can_w4=can_w4
data1$can_mcq4=can_mcq4
data1$can_med4=can_med4
data1$can_cen4=can_cen4
```

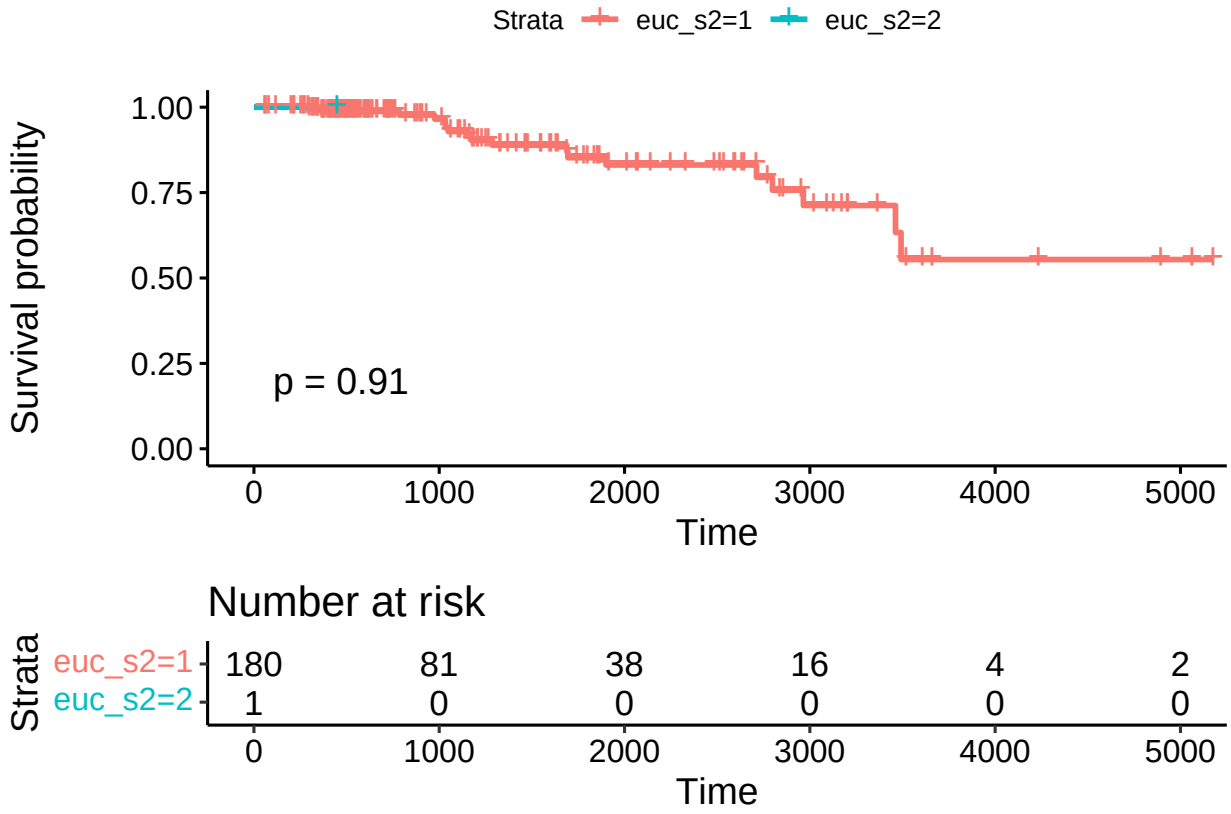
```
HC1 = survfit(obj ~ euc_s2, data = data1)
HC2 = survfit(obj ~ euc_c2, data = data1)
HC3 = survfit(obj ~ euc_a2, data = data1)
HC4 = survfit(obj ~ euc_w2, data = data1)
HC5 = survfit(obj ~ euc_mcq2, data = data1)
HC6 = survfit(obj ~ euc_med2, data = data1)
HC7 = survfit(obj ~ euc_cen2, data = data1)

HC8 = survfit(obj ~ man_s2, data = data1)
HC9 = survfit(obj ~ man_c2, data = data1)
HC10 = survfit(obj ~ man_a2, data = data1)
HC11 = survfit(obj ~ man_w2, data = data1)
HC12 = survfit(obj ~ man_mcq2, data = data1)
HC13 = survfit(obj ~ man_med2, data = data1)
HC14 = survfit(obj ~ man_cen2, data = data1)

HC15 = survfit(obj ~ Jacc_s2, data = data1)
HC16 = survfit(obj ~ Jacc_c2, data = data1)
HC17 = survfit(obj ~ Jacc_a2, data = data1)
HC18 = survfit(obj ~ Jacc_w2, data = data1)
HC19 = survfit(obj ~ Jacc_mcq2, data = data1)
HC20 = survfit(obj ~ Jacc_med2, data = data1)
HC21 = survfit(obj ~ Jacc_cen2, data = data1)

HC22 = survfit(obj ~ max_s2, data = data1)
HC23= survfit(obj ~ max_c2, data = data1)
HC24 = survfit(obj ~ max_a2, data = data1)
HC25 = survfit(obj ~ max_w2, data = data1)
HC26 = survfit(obj ~ max_mcq2, data = data1)
HC27 = survfit(obj ~ max_med2, data = data1)
HC28 = survfit(obj ~ max_cen2, data = data1)

HC29 = survfit(obj ~ bin_s2, data = data1)
HC30 = survfit(obj ~ bin_c2, data = data1)
HC31 = survfit(obj ~ bin_a2, data = data1)
HC32 = survfit(obj ~ bin_w2, data = data1)
HC33 = survfit(obj ~ bin_mcq2, data = data1)
HC34 = survfit(obj ~ bin_med2, data = data1)
HC35 = survfit(obj ~ bin_cen2, data = data1)

HC36 = survfit(obj ~ can_s2, data = data1)
HC37 = survfit(obj ~ can_c2, data = data1)
HC38 = survfit(obj ~ can_a2, data = data1)
HC39 = survfit(obj ~ can_w2, data = data1)
HC40 = survfit(obj ~ can_mcq2, data = data1)
```

```
HC41 = survfit(obj ~ can_med2, data = data1)
HC42 = survfit(obj ~ can_cen2, data = data1)

HW1 = survfit(obj ~ euc_s4, data = data1)
HW2 = survfit(obj ~ euc_c4, data = data1)
HW3 = survfit(obj ~ euc_a4, data = data1)
HW4 = survfit(obj ~ euc_w4, data = data1)
HW5 = survfit(obj ~ euc_mcq4, data = data1)
HW6 = survfit(obj ~ euc_med4, data = data1)
HW7 = survfit(obj ~ euc_cen4, data = data1)

HW8 = survfit(obj ~ man_s4, data = data1)
HW9 = survfit(obj ~ man_c4, data = data1)
HW10 = survfit(obj ~ man_a4, data = data1)
HW11 = survfit(obj ~ man_w4, data = data1)
HW12 = survfit(obj ~ man_mcq4, data = data1)
HW13 = survfit(obj ~ man_med4, data = data1)
HW14 = survfit(obj ~ man_cen4, data = data1)

HW15 = survfit(obj ~ Jacc_s4, data = data1)
HW16 = survfit(obj ~ Jacc_c4, data = data1)
HW17 = survfit(obj ~ Jacc_a4, data = data1)
HW18 = survfit(obj ~ Jacc_w4, data = data1)
HW19 = survfit(obj ~ Jacc_mcq4, data = data1)
HW20 = survfit(obj ~ Jacc_med4, data = data1)
HW21 = survfit(obj ~ Jacc_cen4, data = data1)

HW22 = survfit(obj ~ max_s4, data = data1)
HW23= survfit(obj ~ max_c4, data = data1)
HW24 = survfit(obj ~ max_a4, data = data1)
HW25 = survfit(obj ~ max_w4, data = data1)
HW26 = survfit(obj ~ max_mcq4, data = data1)
HW27 = survfit(obj ~ max_med4, data = data1)
HW28 = survfit(obj ~ max_cen4, data = data1)

HW29 = survfit(obj ~ bin_s4, data = data1)
HW30 = survfit(obj ~ bin_c4, data = data1)
HW31 = survfit(obj ~ bin_a4, data = data1)
HW32 = survfit(obj ~ bin_w4, data = data1)
HW33 = survfit(obj ~ bin_mcq4, data = data1)
HW34 = survfit(obj ~ bin_med4, data = data1)
HW35 = survfit(obj ~ bin_cen4, data = data1)

HW36 = survfit(obj ~ can_s4, data = data1)
HW37 = survfit(obj ~ can_c4, data = data1)
HW38 = survfit(obj ~ can_a4, data = data1)
HW39 = survfit(obj ~ can_w4, data = data1)
HW40 = survfit(obj ~ can_mcq4, data = data1)
HW41 = survfit(obj ~ can_med4, data = data1)
HW42 = survfit(obj ~ can_cen4, data = data1)


HS1 = survfit(obj ~ euc_s3, data = data1)
```

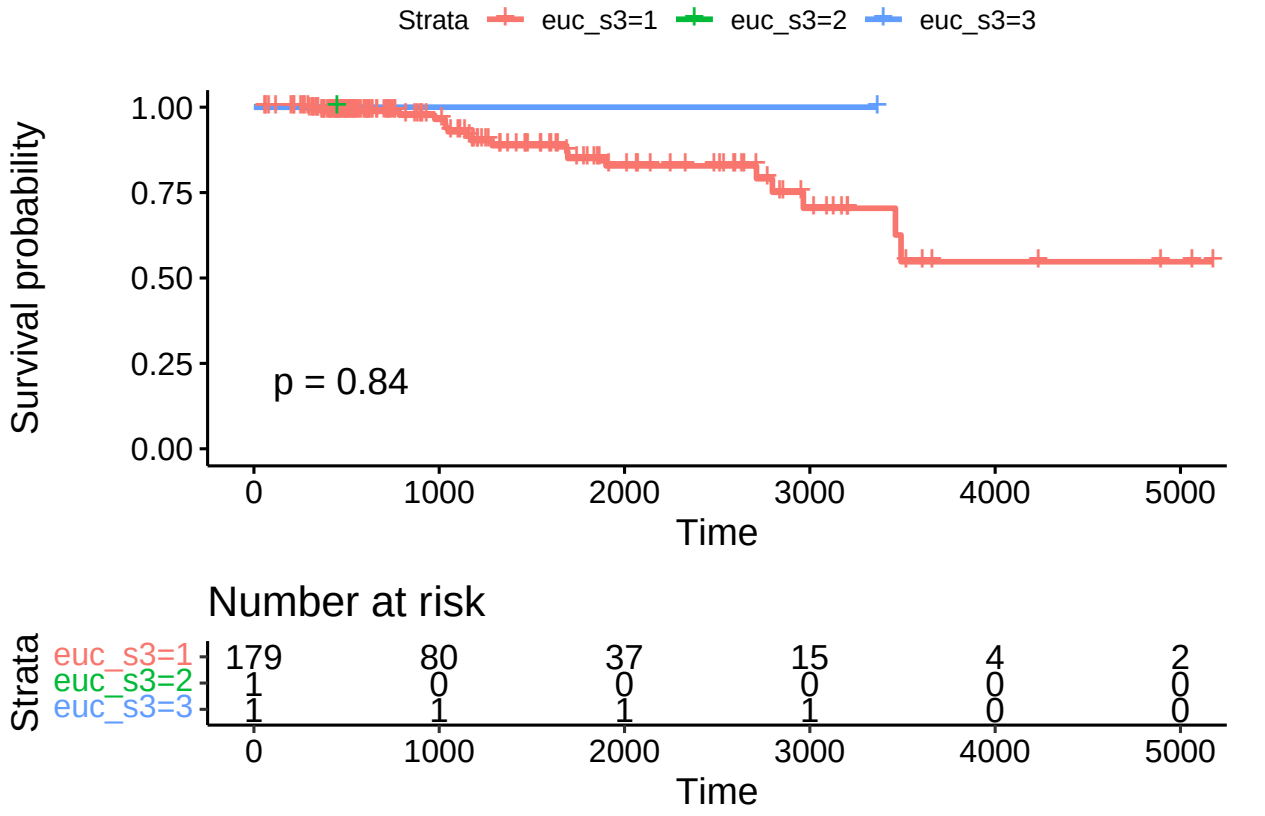
```
HS2 = survfit(obj ~ euc_c3, data = data1)
HS3 = survfit(obj ~ euc_a3, data = data1)
HS4 = survfit(obj ~ euc_w3, data = data1)
HS5 = survfit(obj ~ euc_mcq3, data = data1)
HS6 = survfit(obj ~ euc_med3, data = data1)
HS7 = survfit(obj ~ euc_cen3, data = data1)

HS8 = survfit(obj ~ man_s3, data = data1)
HS9 = survfit(obj ~ man_c3, data = data1)
HS10 = survfit(obj ~ man_a3, data = data1)
HS11 = survfit(obj ~ man_w3, data = data1)
HS12 = survfit(obj ~ man_mcq3, data = data1)
HS13 = survfit(obj ~ man_med3, data = data1)
HS14 = survfit(obj ~ man_cen3, data = data1)

HS15 = survfit(obj ~ Jacc_s3, data = data1)
HS16 = survfit(obj ~ Jacc_c3, data = data1)
HS17 = survfit(obj ~ Jacc_a3, data = data1)
HS18 = survfit(obj ~ Jacc_w3, data = data1)
HS19 = survfit(obj ~ Jacc_mcq3, data = data1)
HS20 = survfit(obj ~ Jacc_med3, data = data1)
HS21 = survfit(obj ~ Jacc_cen3, data = data1)

HS22 = survfit(obj ~ max_s3, data = data1)
HS23= survfit(obj ~ max_c3, data = data1)
HS24 = survfit(obj ~ max_a3, data = data1)
HS25 = survfit(obj ~ max_w3, data = data1)
HS26 = survfit(obj ~ max_mcq3, data = data1)
HS27 = survfit(obj ~ max_med3, data = data1)
HS28 = survfit(obj ~ max_cen3, data = data1)

HS29 = survfit(obj ~ bin_s3, data = data1)
HS30 = survfit(obj ~ bin_c3, data = data1)
HS31 = survfit(obj ~ bin_a3, data = data1)
HS32 = survfit(obj ~ bin_w3, data = data1)
HS33 = survfit(obj ~ bin_mcq3, data = data1)
HS34 = survfit(obj ~ bin_med3, data = data1)
HS35 = survfit(obj ~ bin_cen3, data = data1)

HS36 = survfit(obj ~ can_s3, data = data1)
HS37 = survfit(obj ~ can_c3, data = data1)
HS38 = survfit(obj ~ can_a3, data = data1)
HS39 = survfit(obj ~ can_w3, data = data1)
HS40 = survfit(obj ~ can_mcq3, data = data1)
HS41 = survfit(obj ~ can_med3, data = data1)
HS42 = survfit(obj ~ can_cen3, data = data1)
```

4.2.Fit the results of Hierarchical clustering 4.2.1. Euclidean distance measurement + Single Linkage

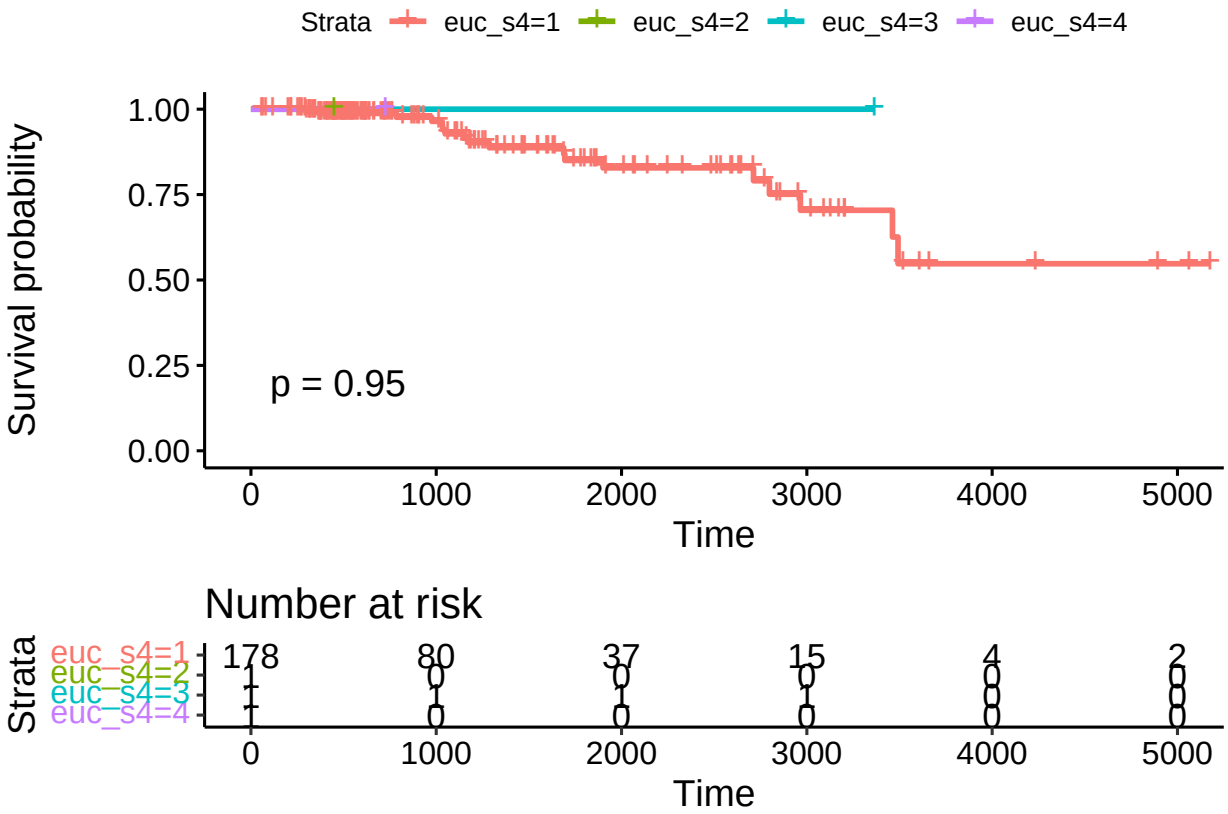
```
ggsurvplot(HC1,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS1,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

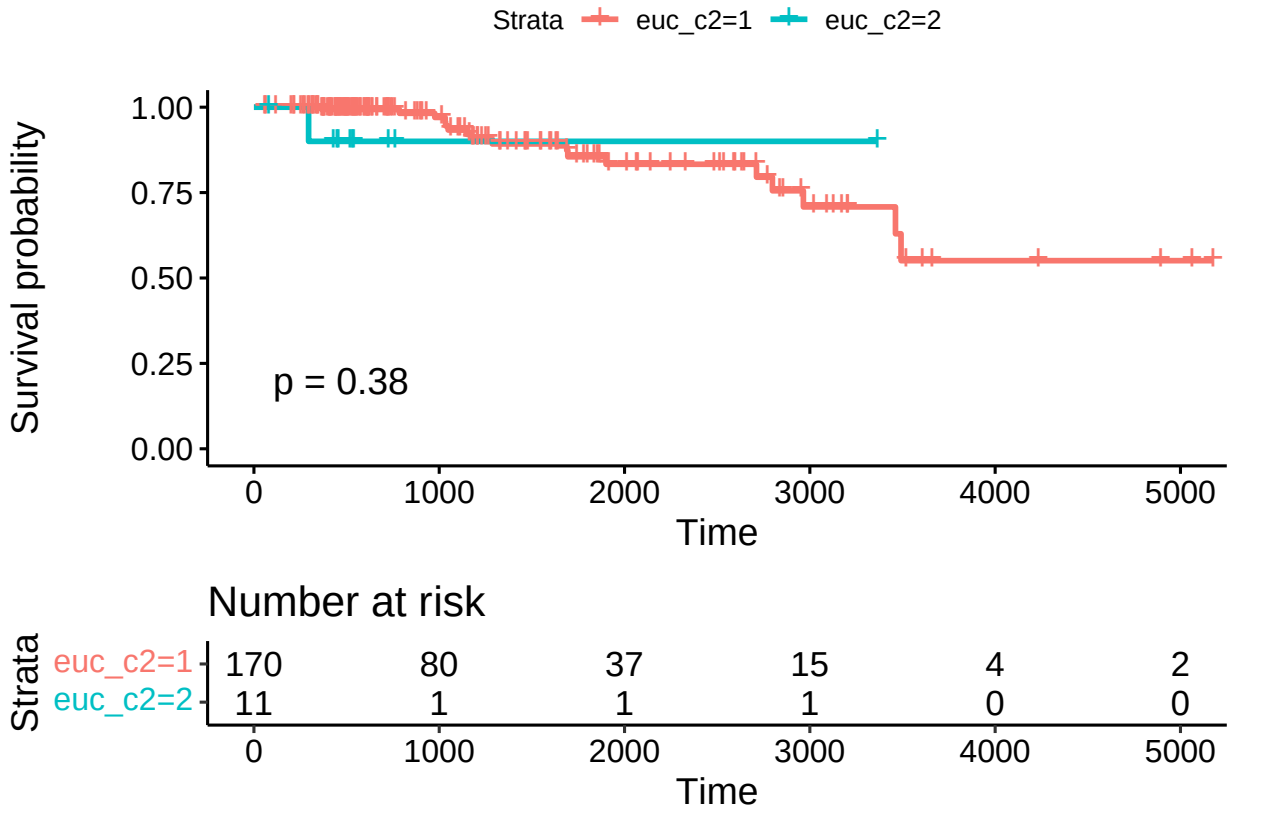



```
ggsurvplot(HW1,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

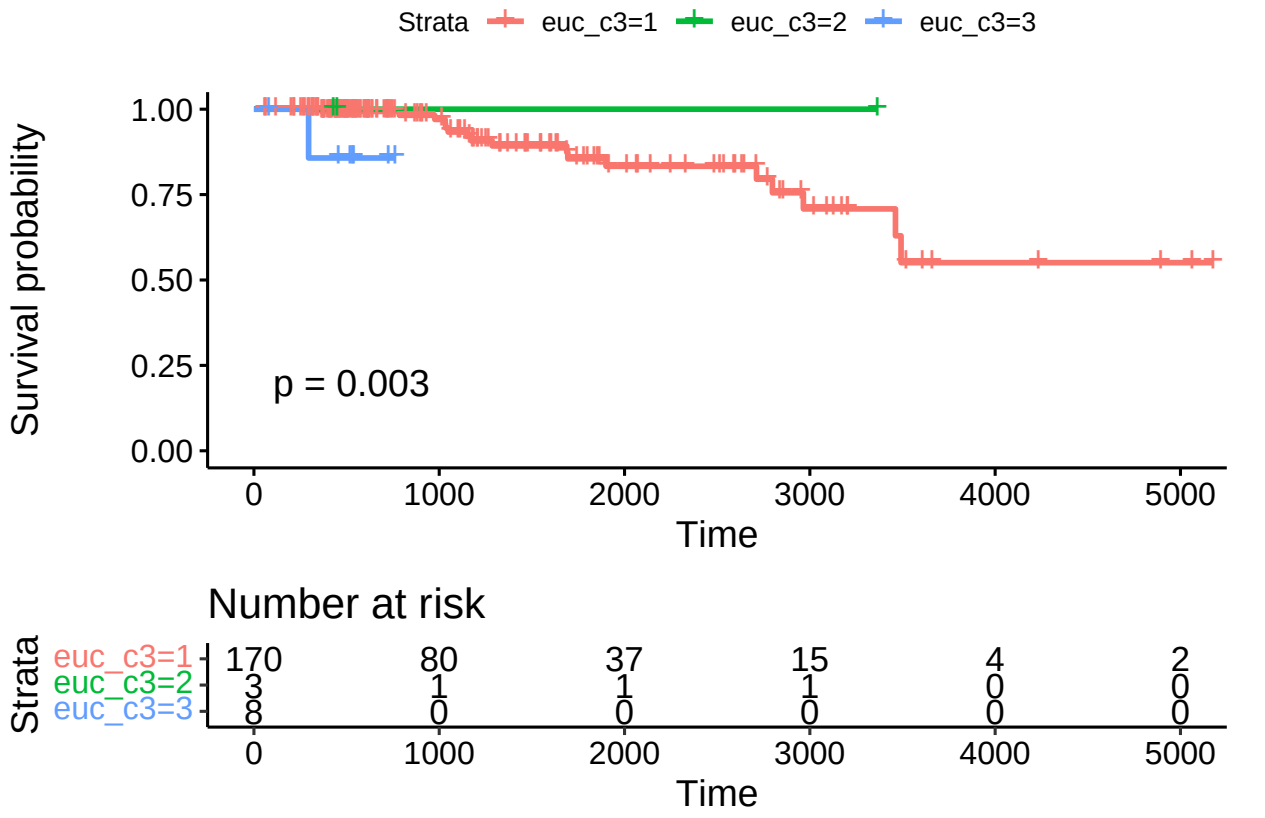


4.2.2. Euclidean distance measurement + Average Linkage

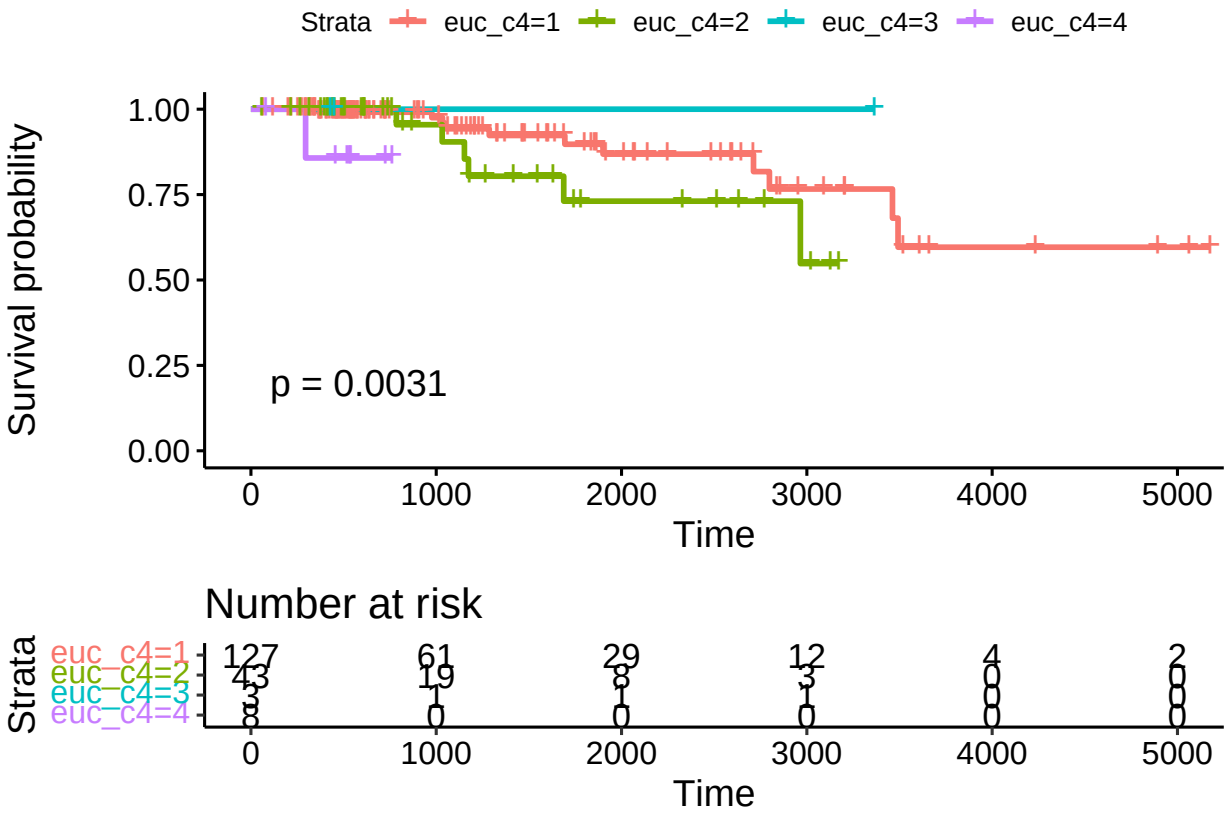
```
ggsurvplot(HC2,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS2,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

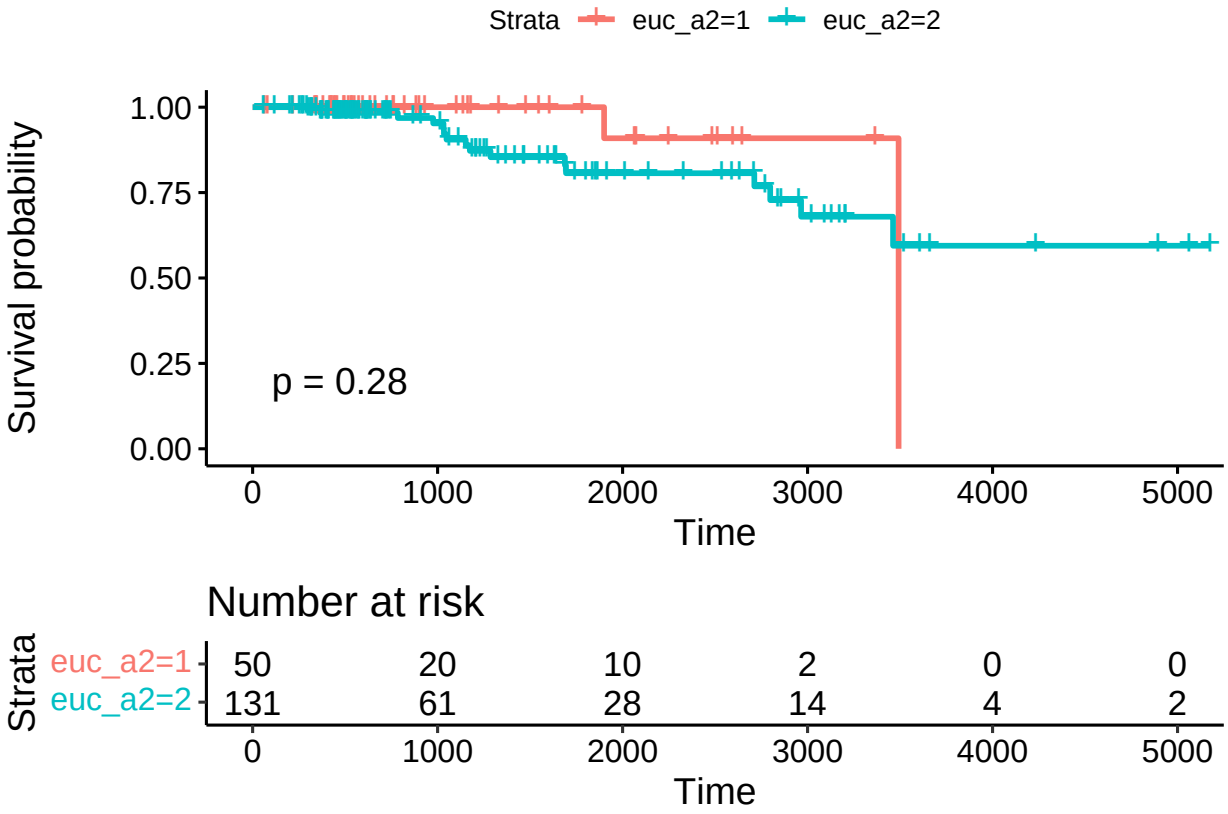



```
ggsurvplot(HW2,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

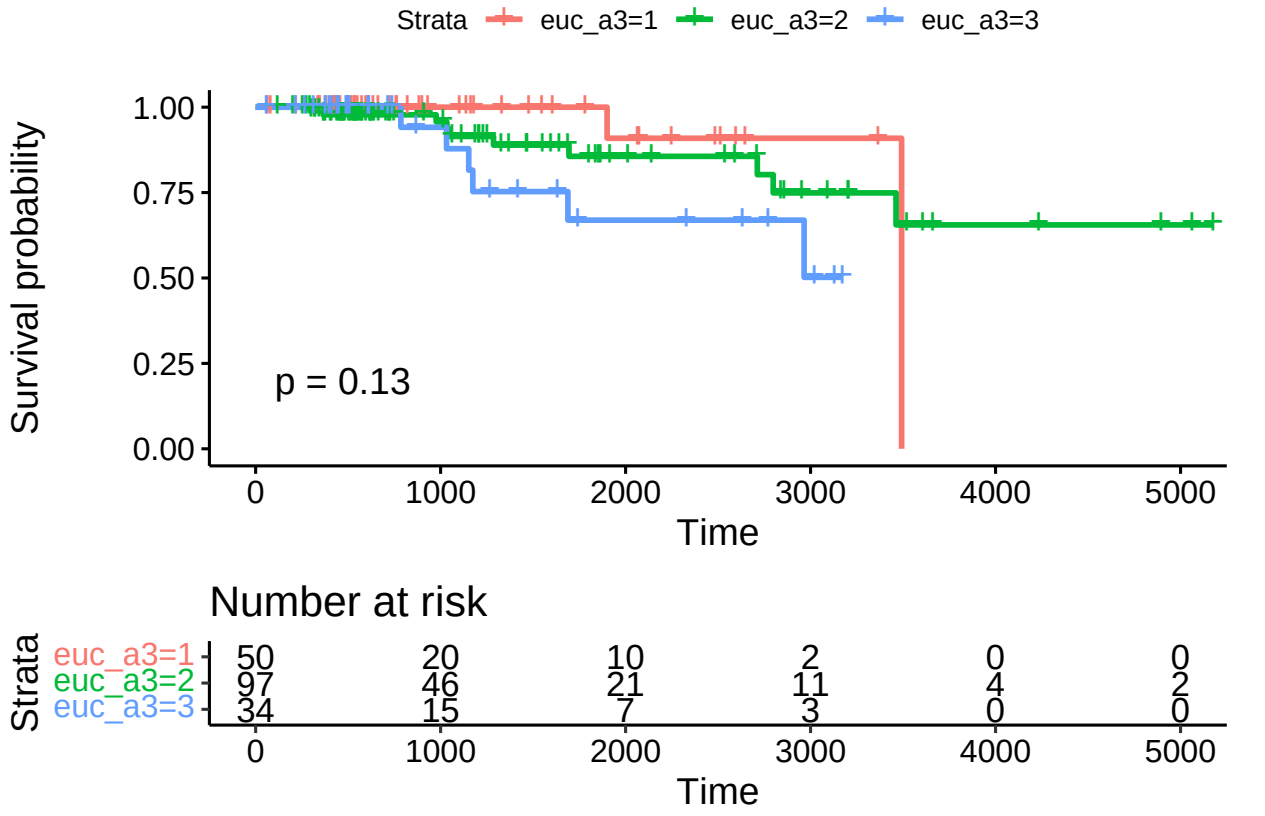


4.2.3. Euclidean distance measurement + Complete Linkage

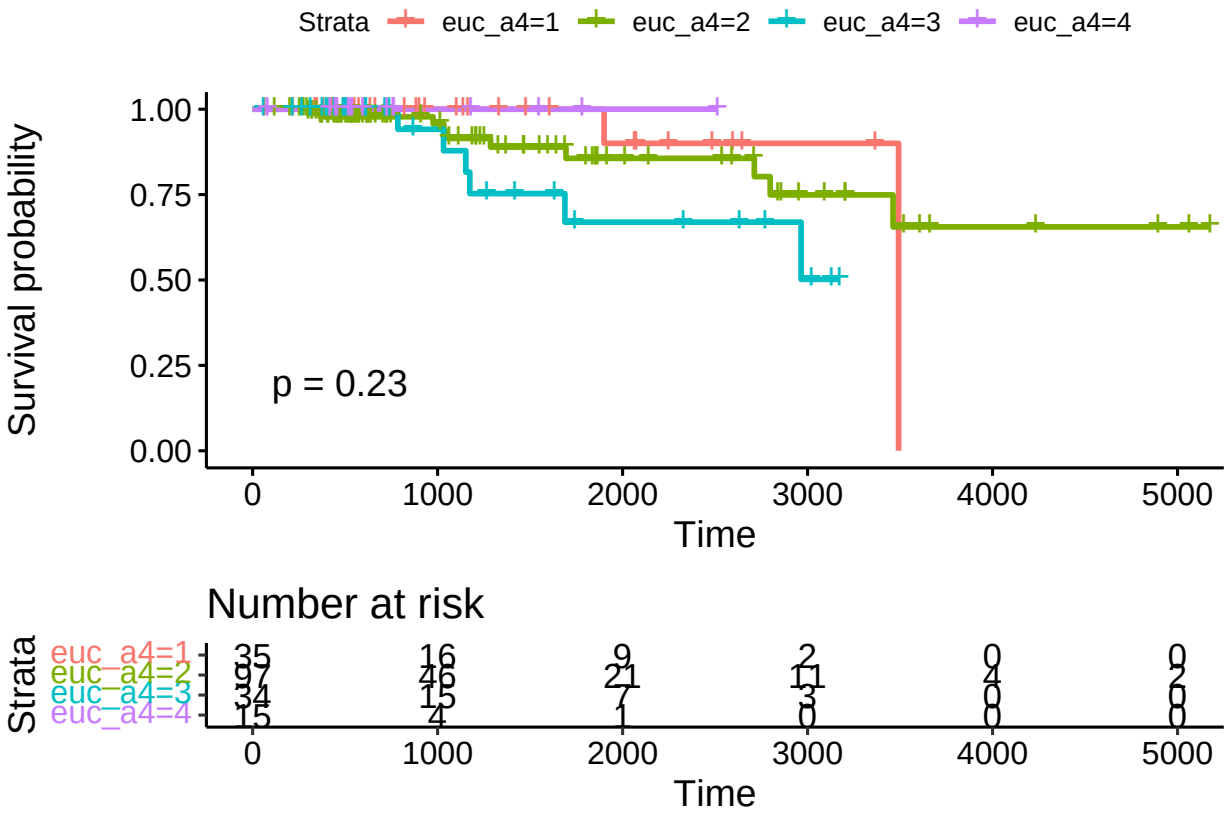
```
ggsurvplot(HC3,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS3,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

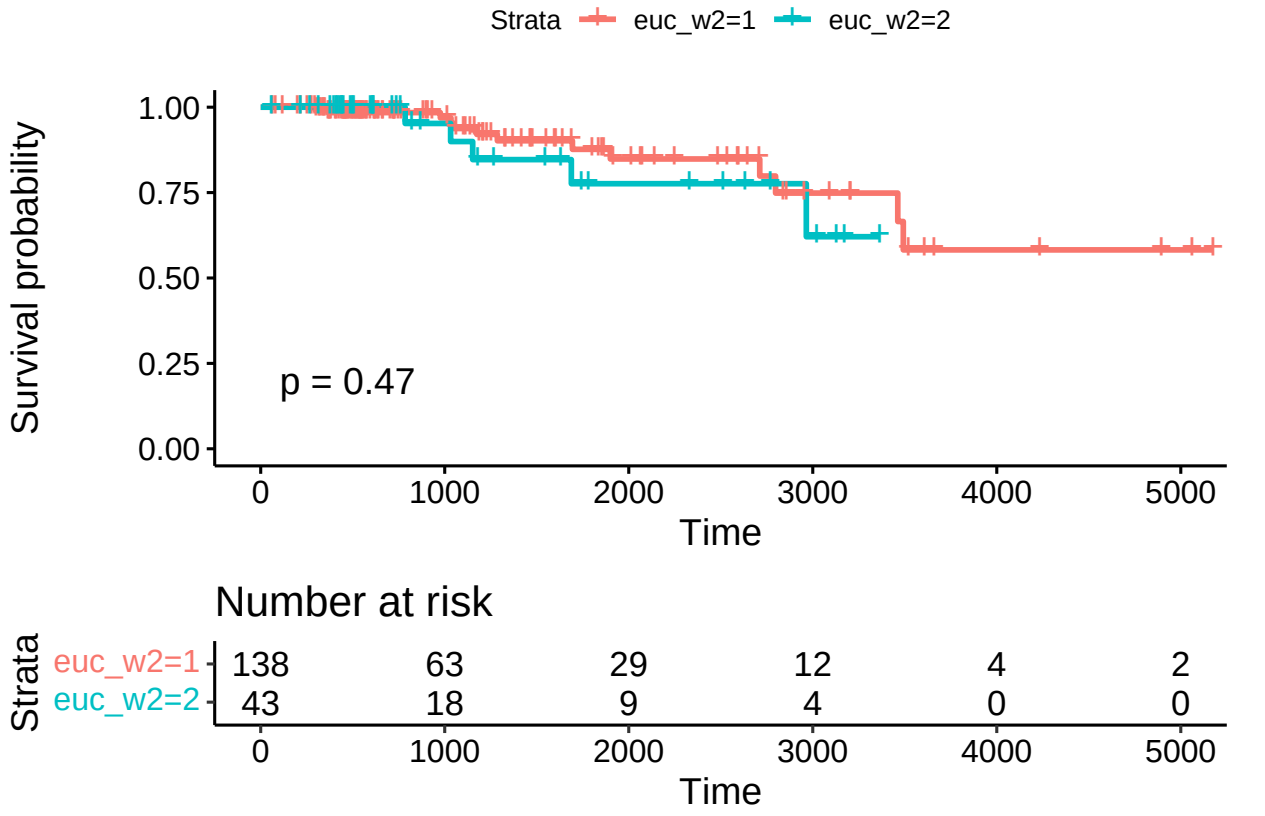



```
ggsurvplot(HW3,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

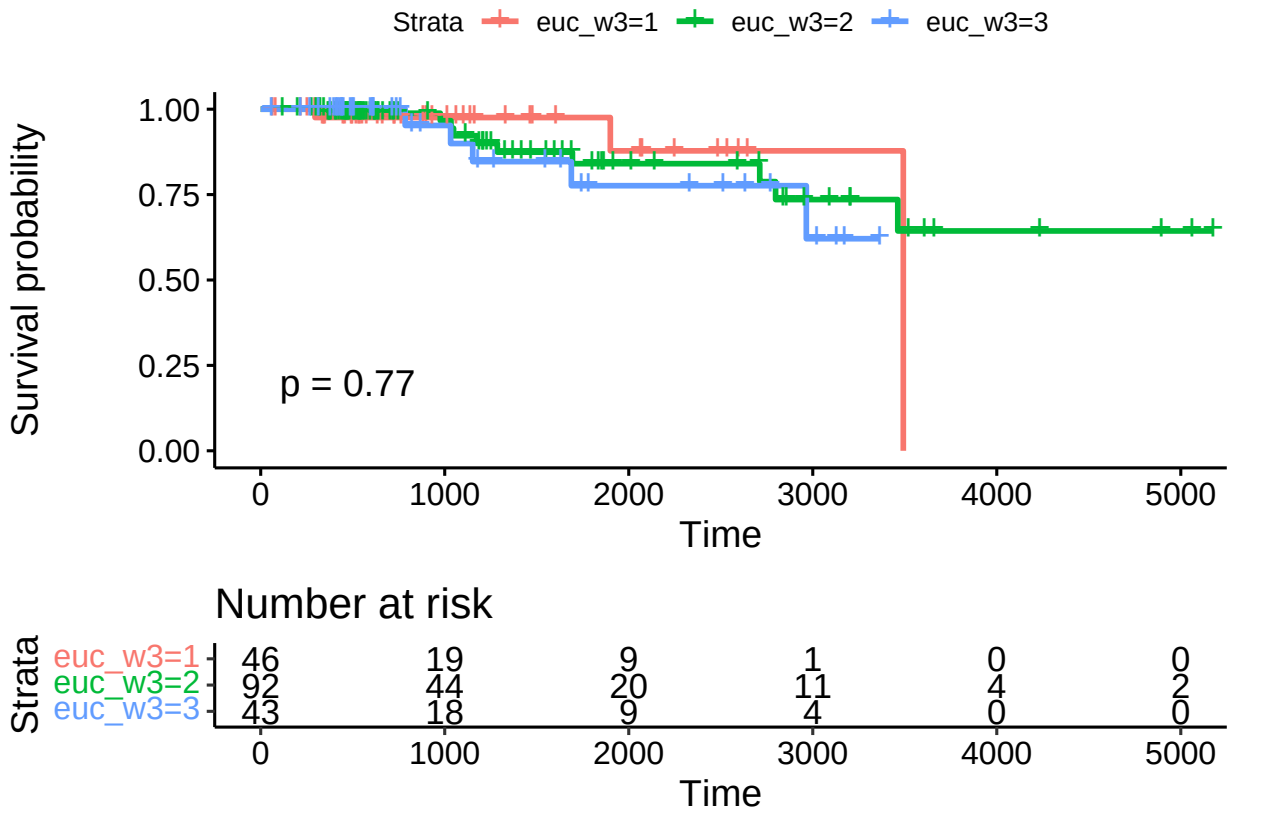


4.2.4. Euclidean distance measurement + Ward Linkage

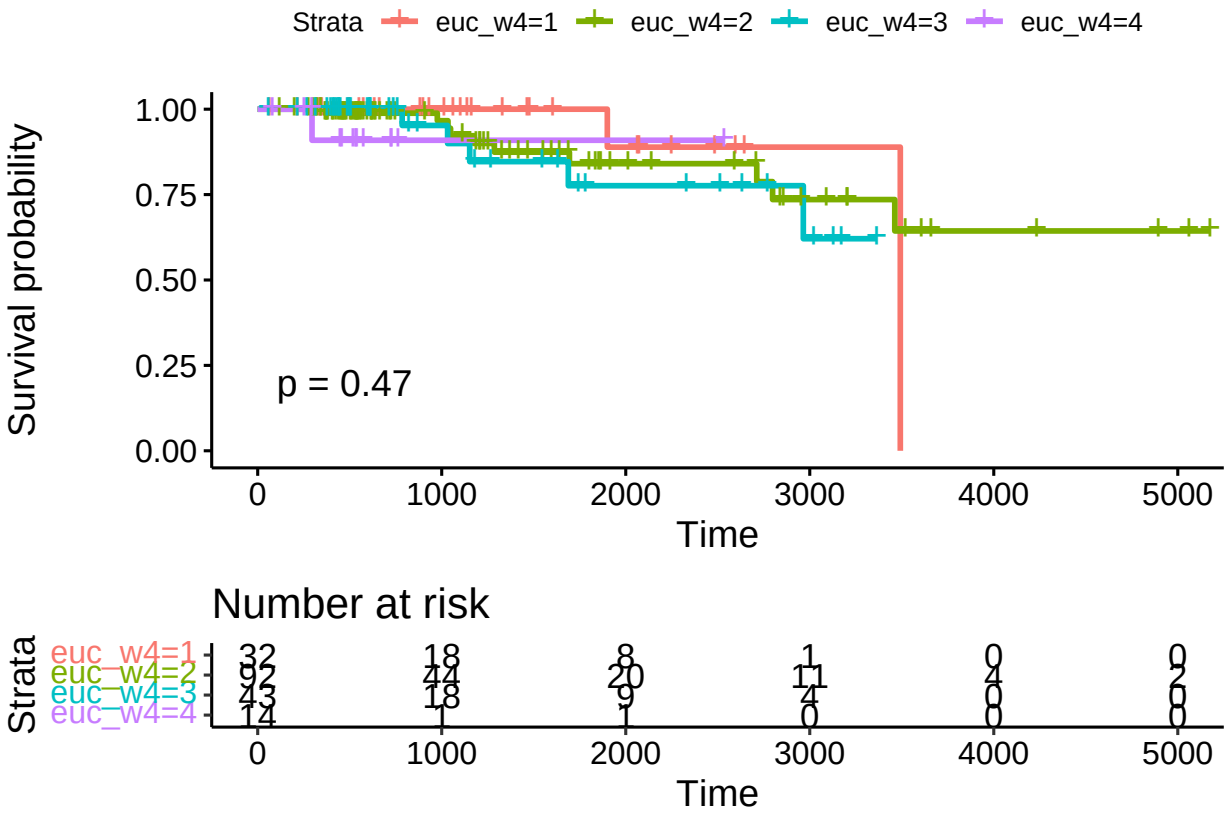
```
ggsurvplot(HC4,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS4,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

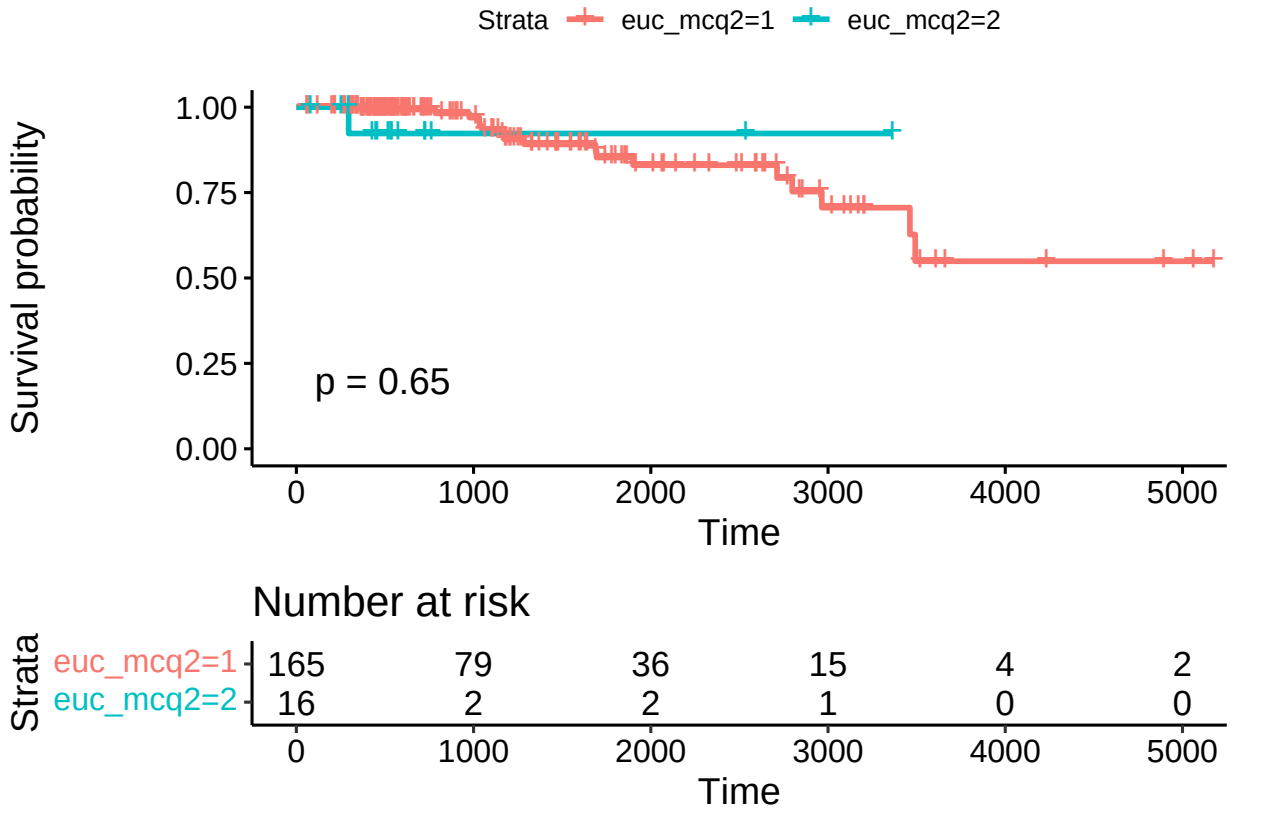



```
ggsurvplot(HW4,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

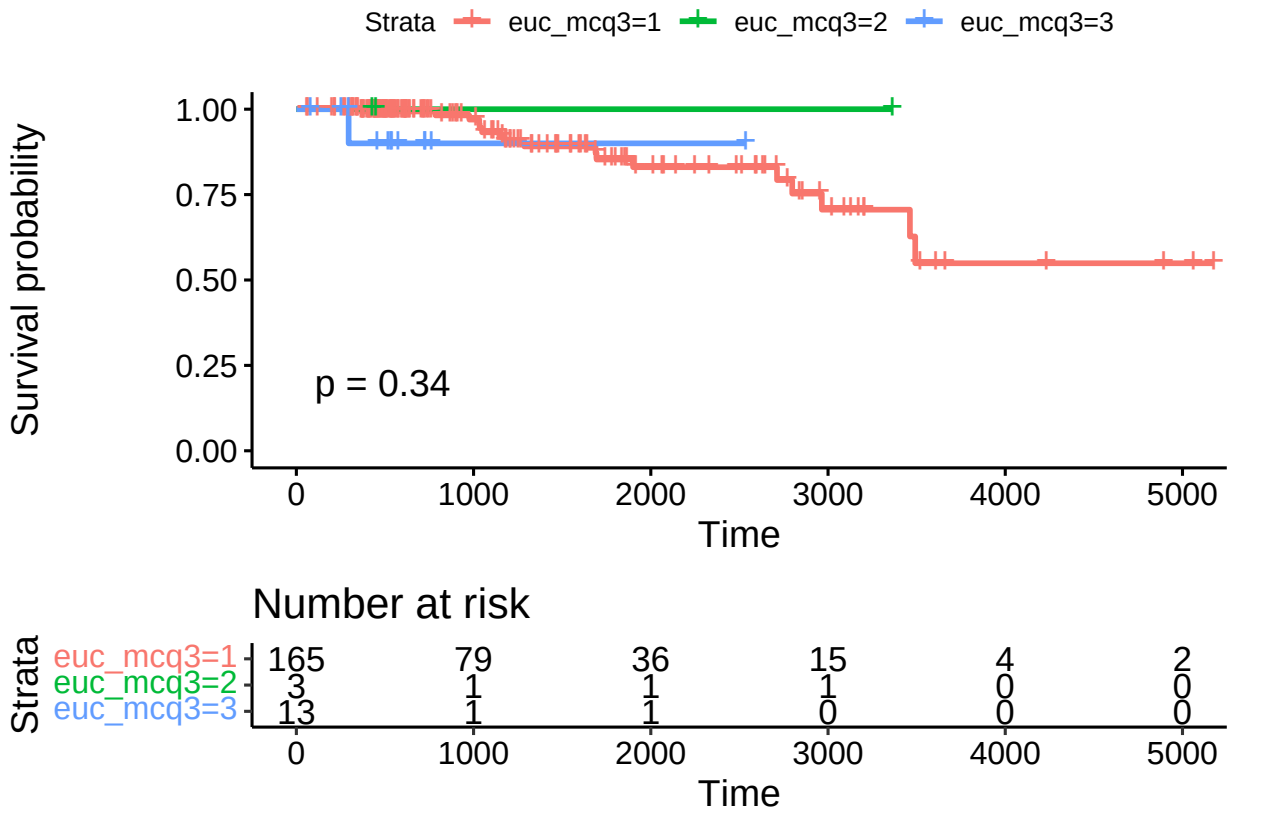


4.2.5. Euclidean distance measurement + Mcquitty Linkage

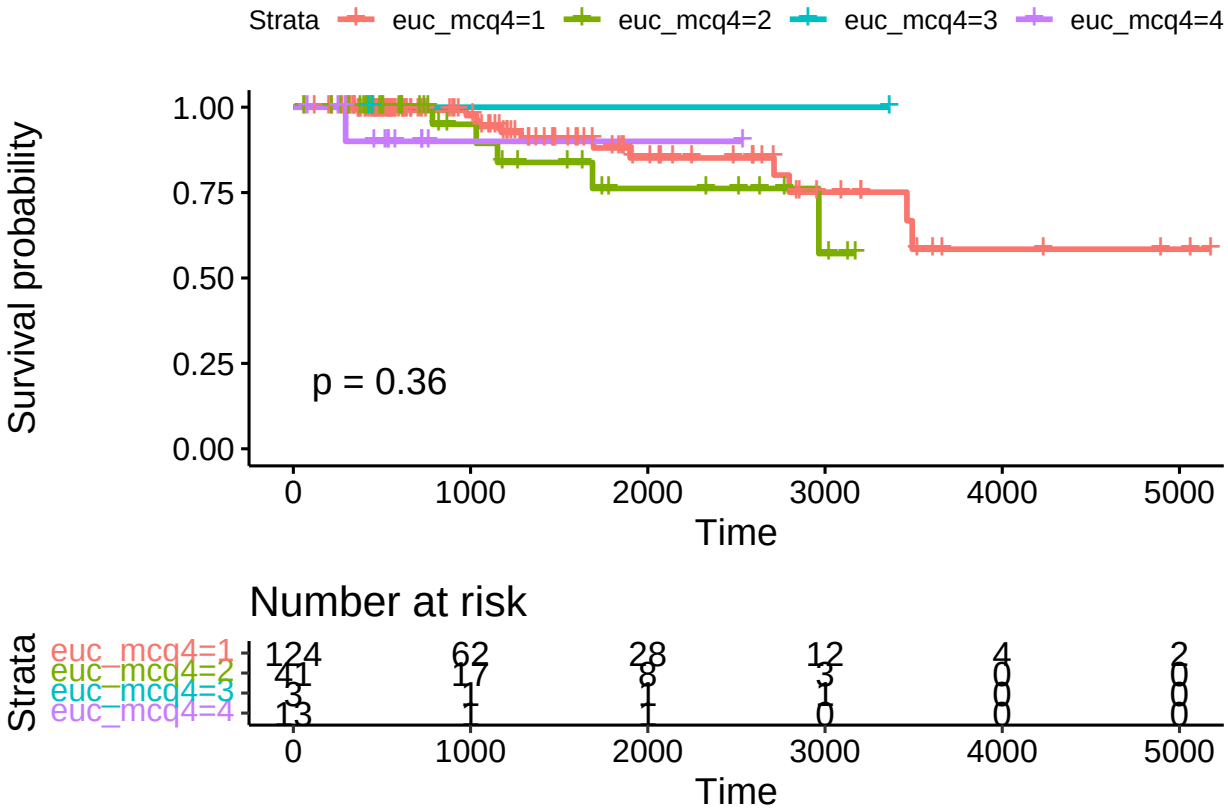
```
ggsurvplot(HC5,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS5,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

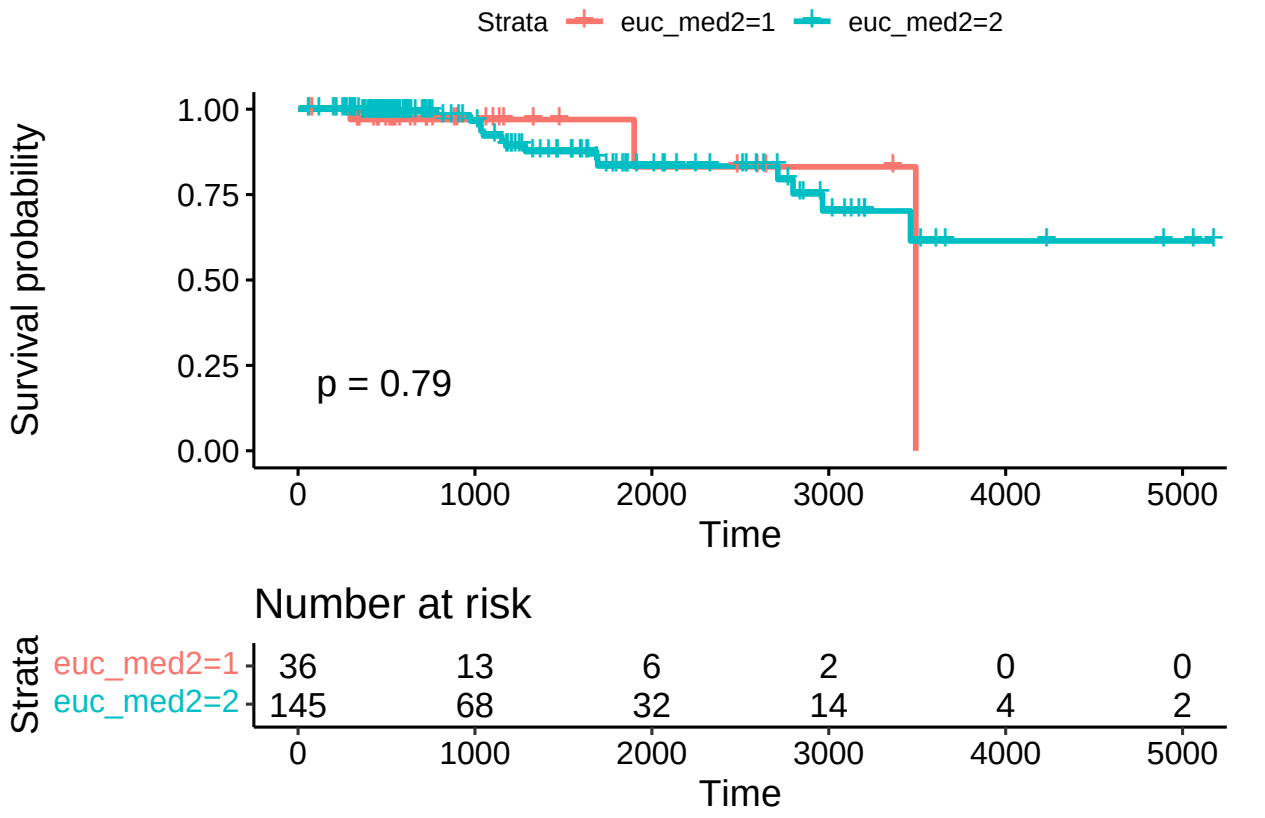



```
ggsurvplot(HW5,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

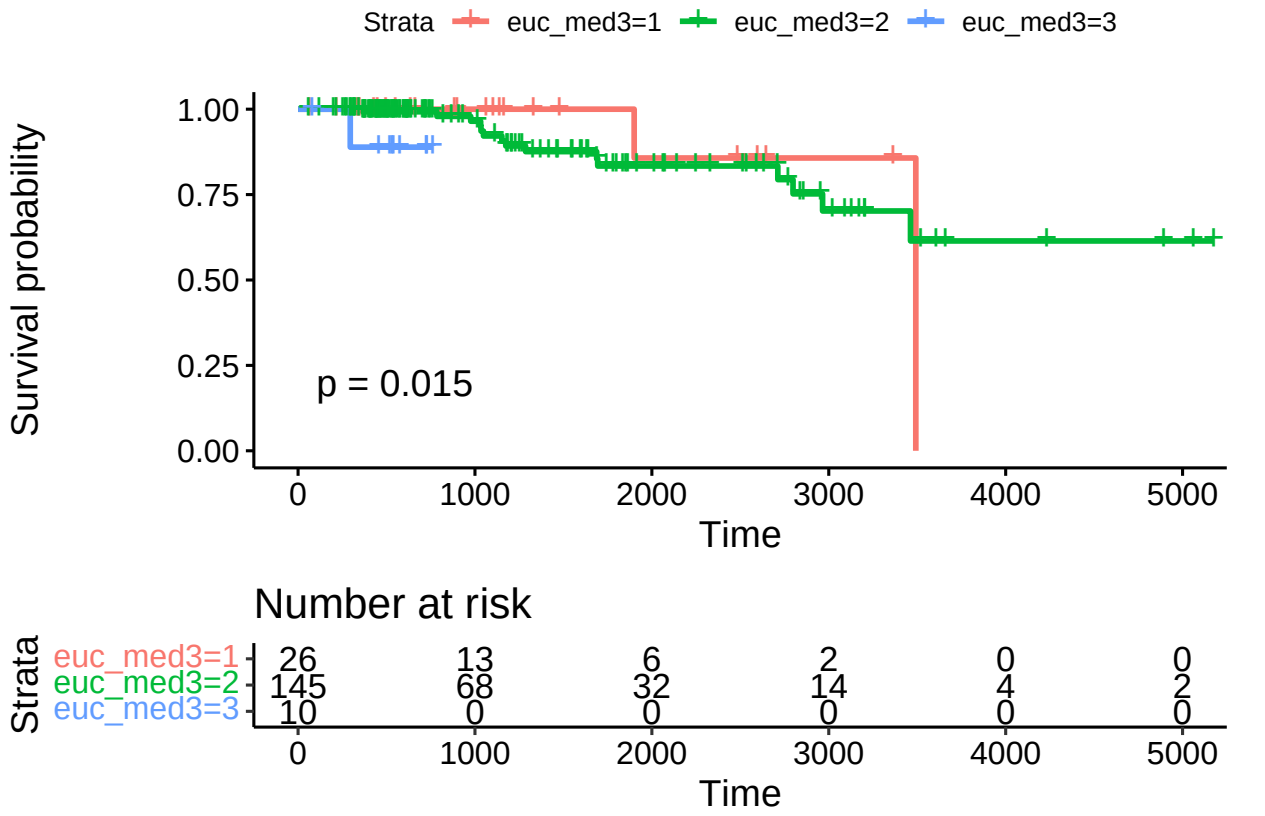


4.2.6. Euclidean distance measurement + Median Linkage

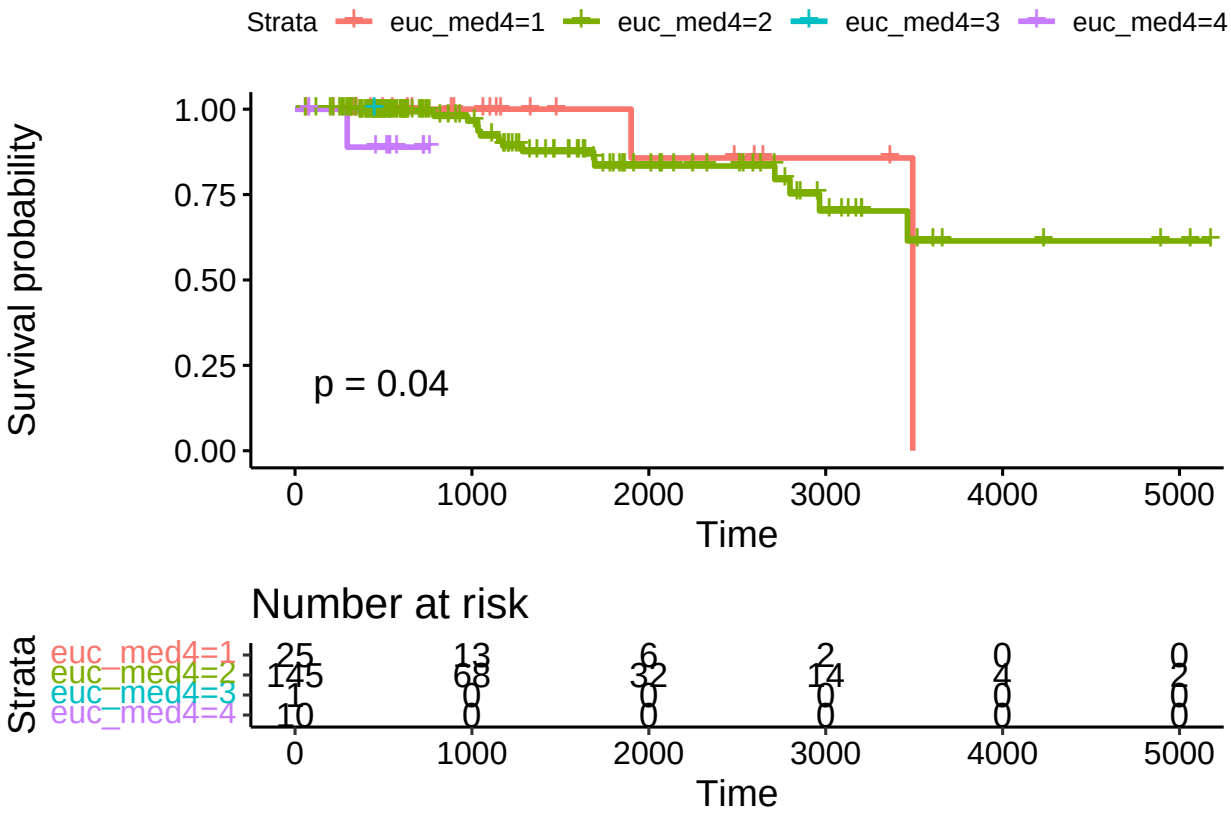
```
ggsurvplot(HC6,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS6,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

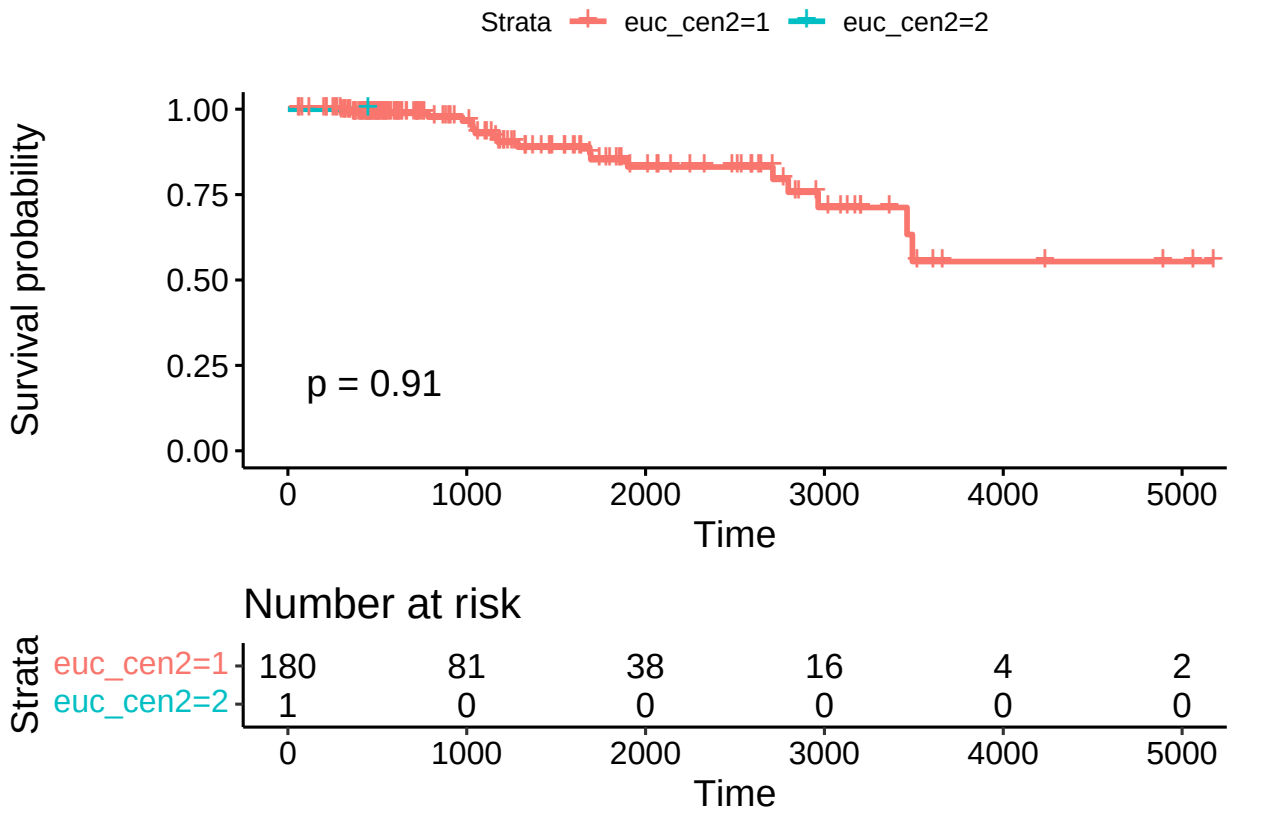



```
ggsurvplot(HW6,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

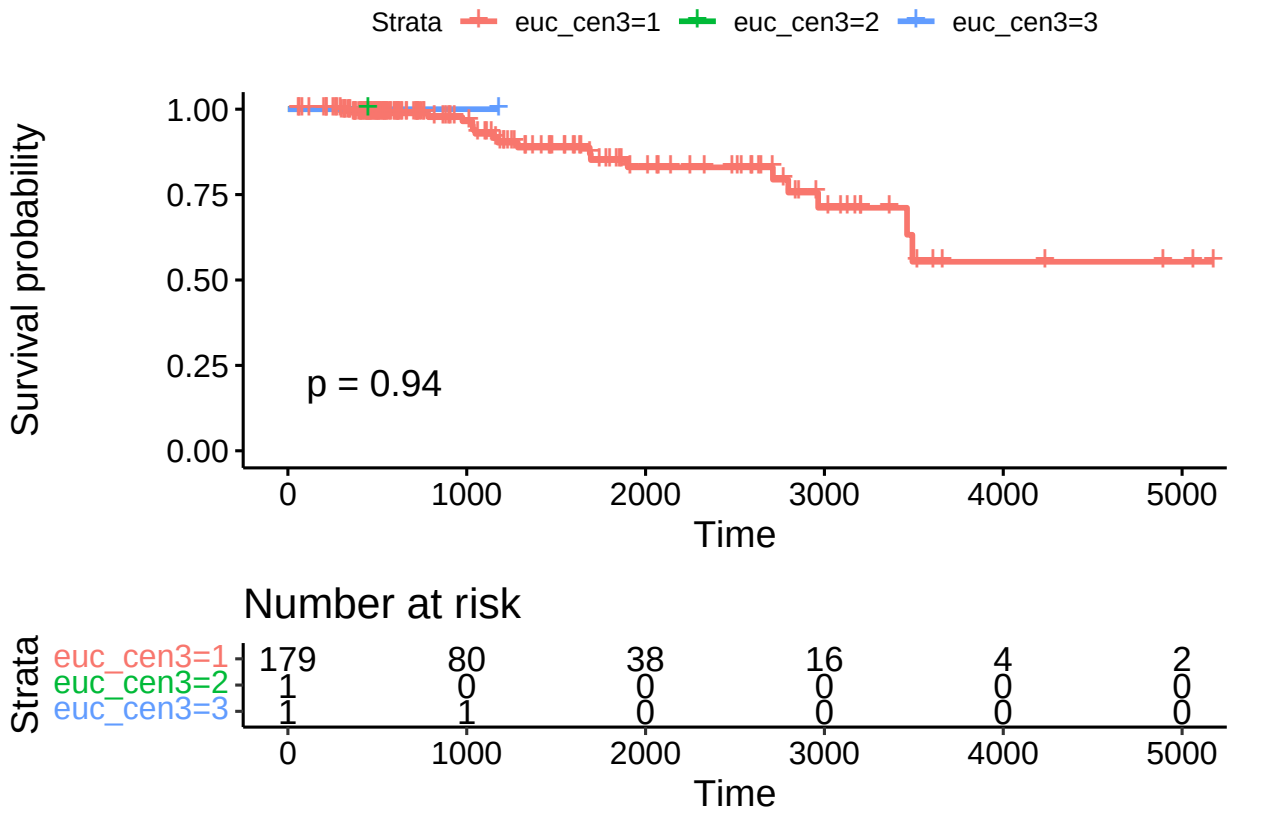


4.2.7. Euclidean distance measurement + Centroid Linkage

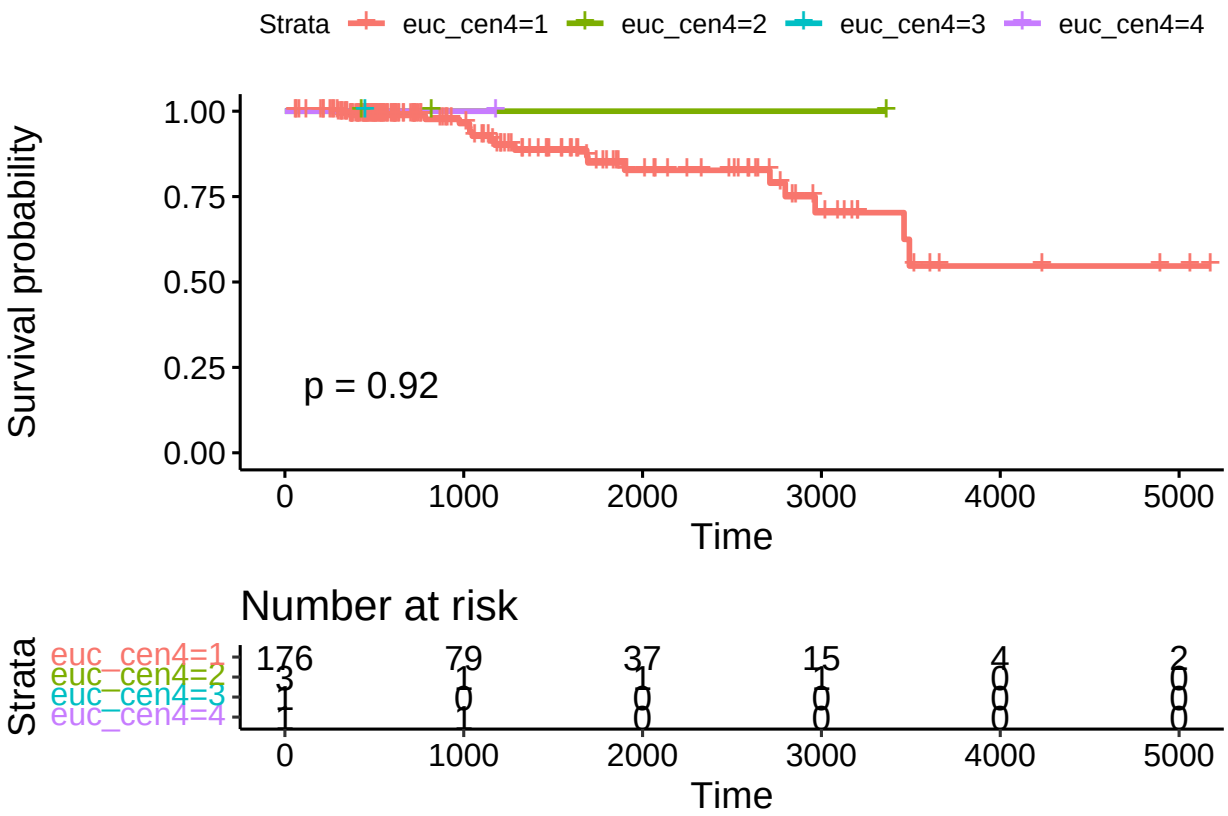
```
ggsurvplot(HC7,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS7,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

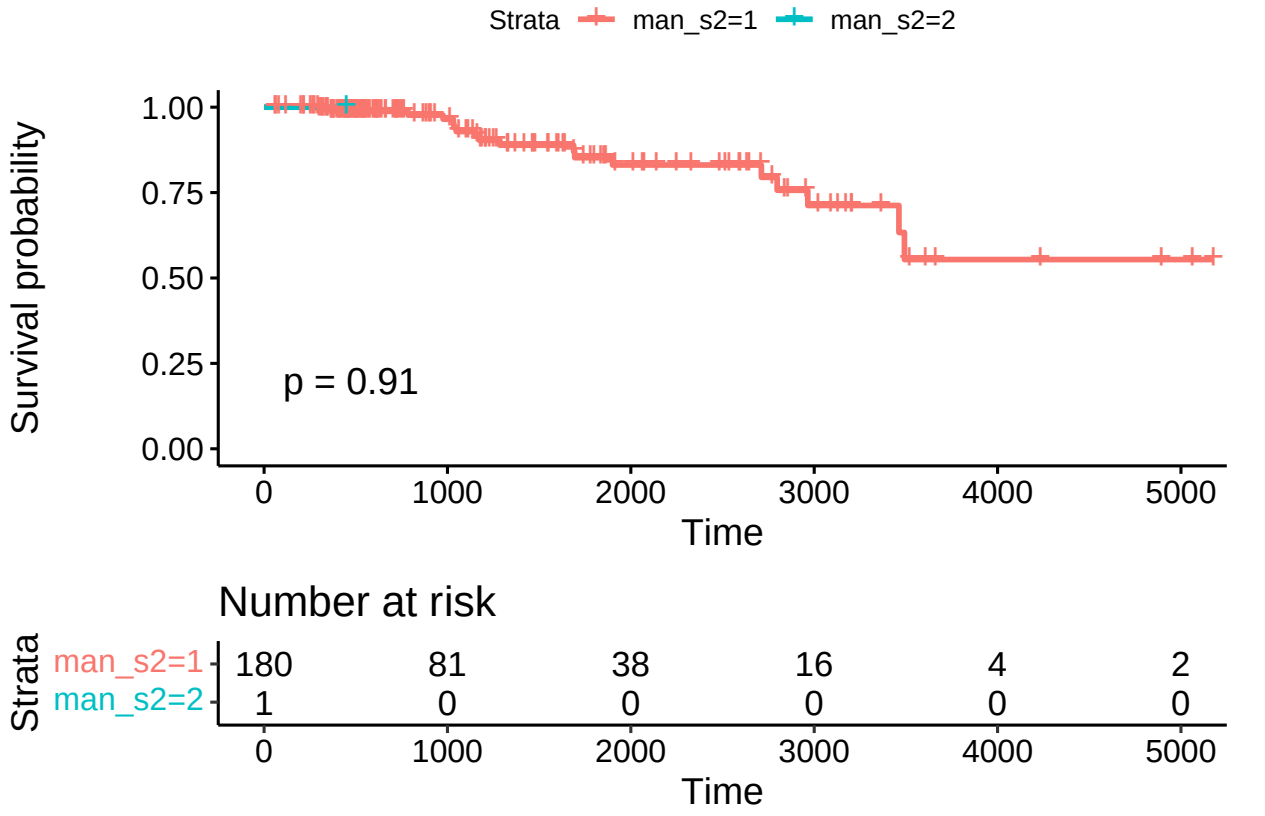



```
ggsurvplot(HW7,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

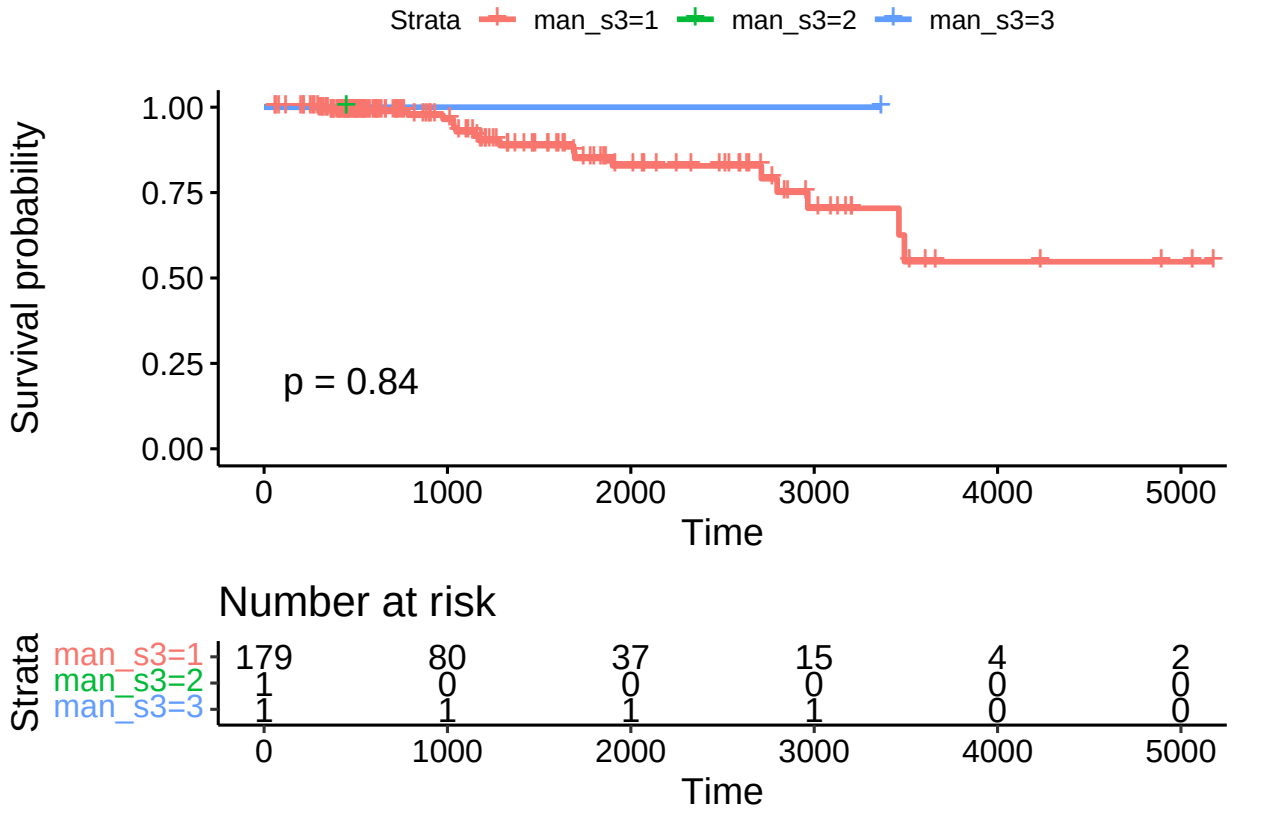


4.2.8. Manhattan distance measurement + Single Linkage

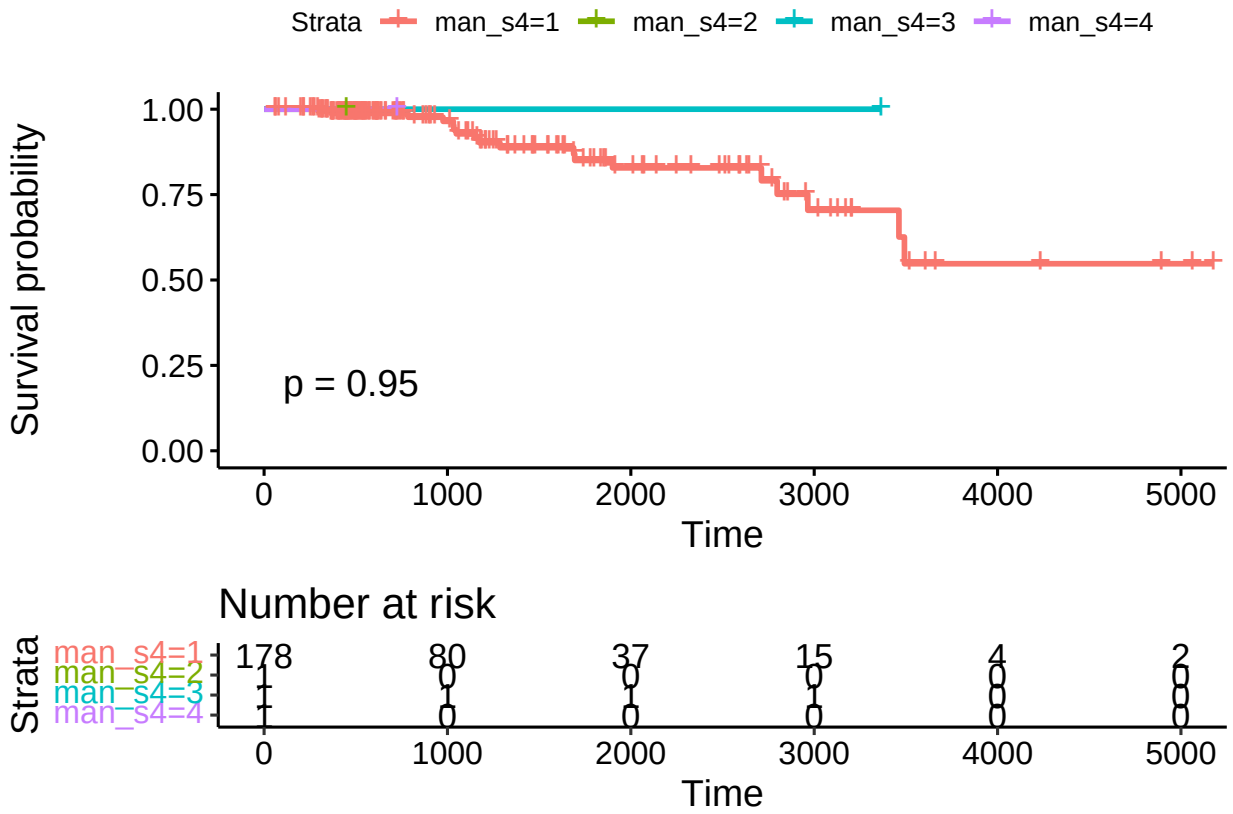
```
ggsurvplot(HC8,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS8,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

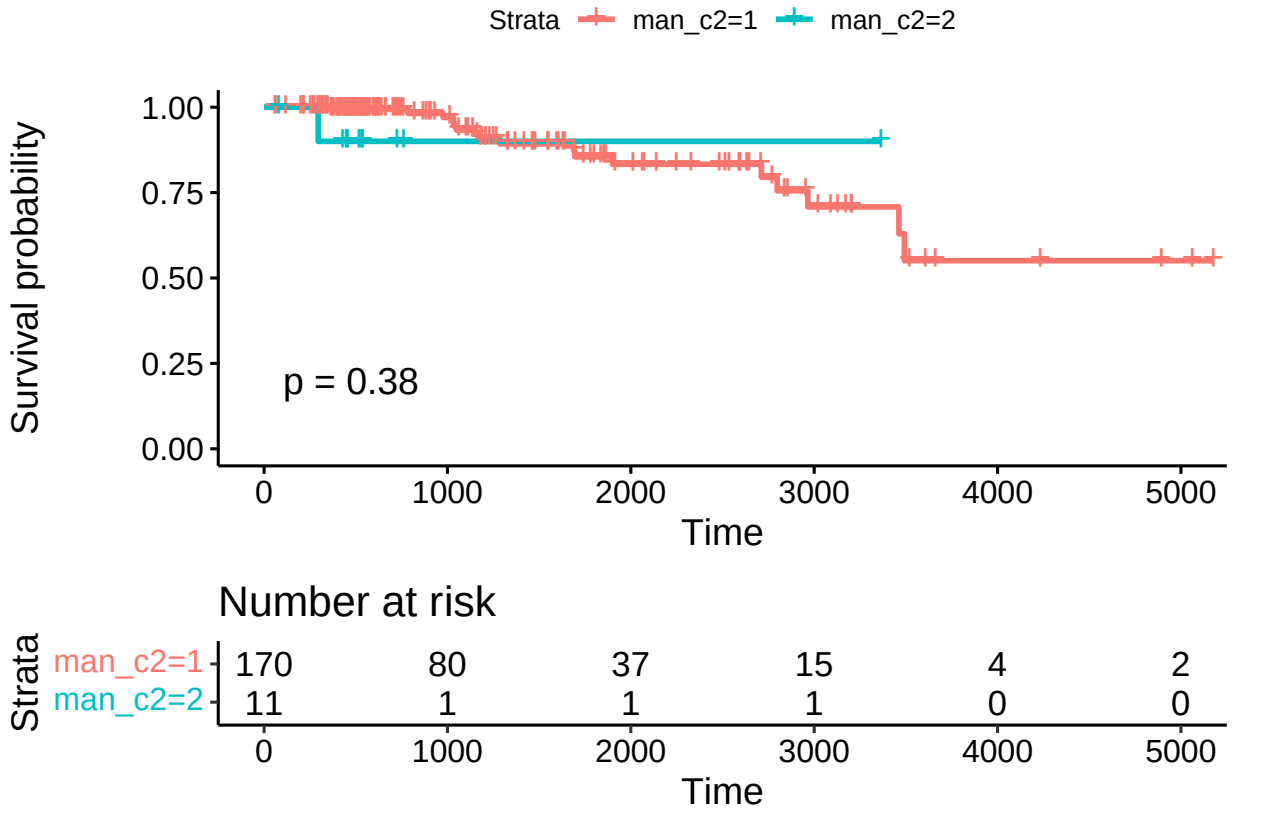



```
ggsurvplot(HW8,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

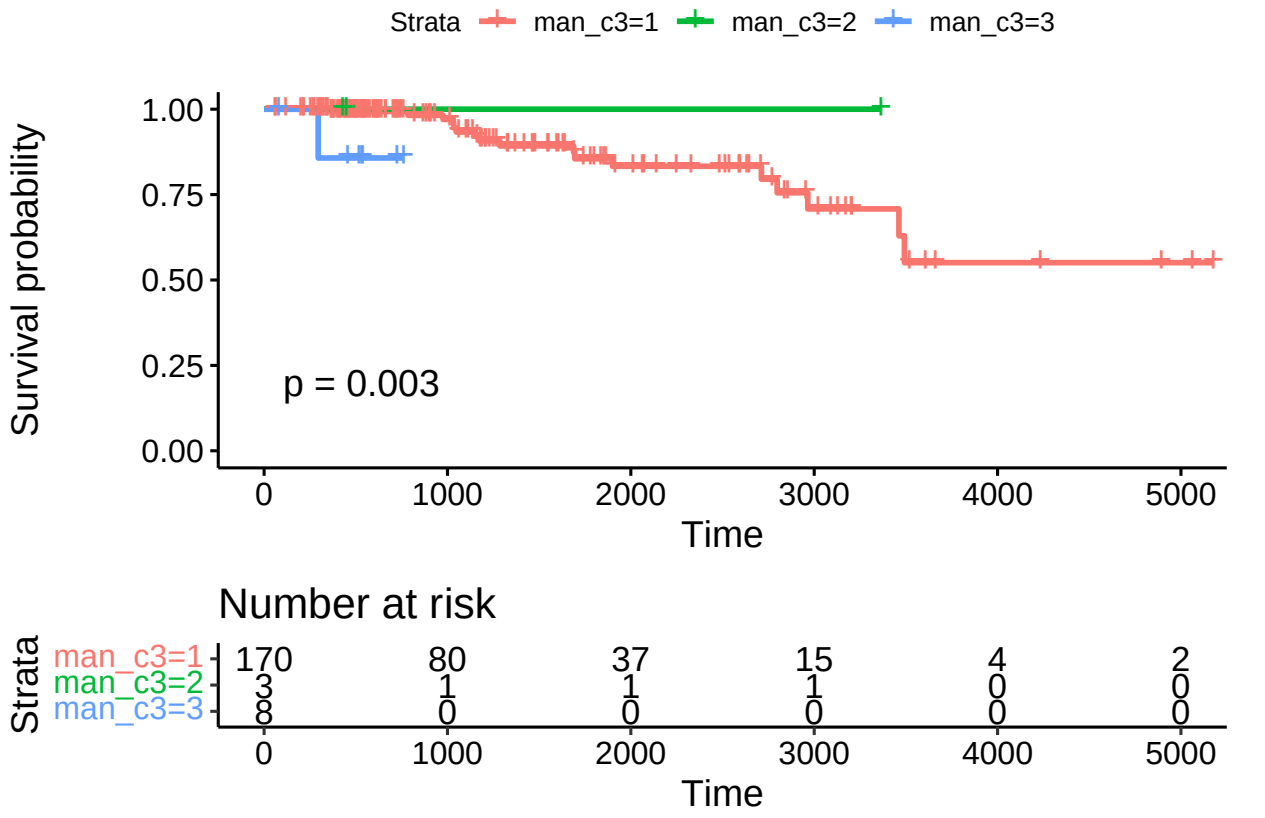


4.2.9. Manhattan distance measurement + Average Linkage

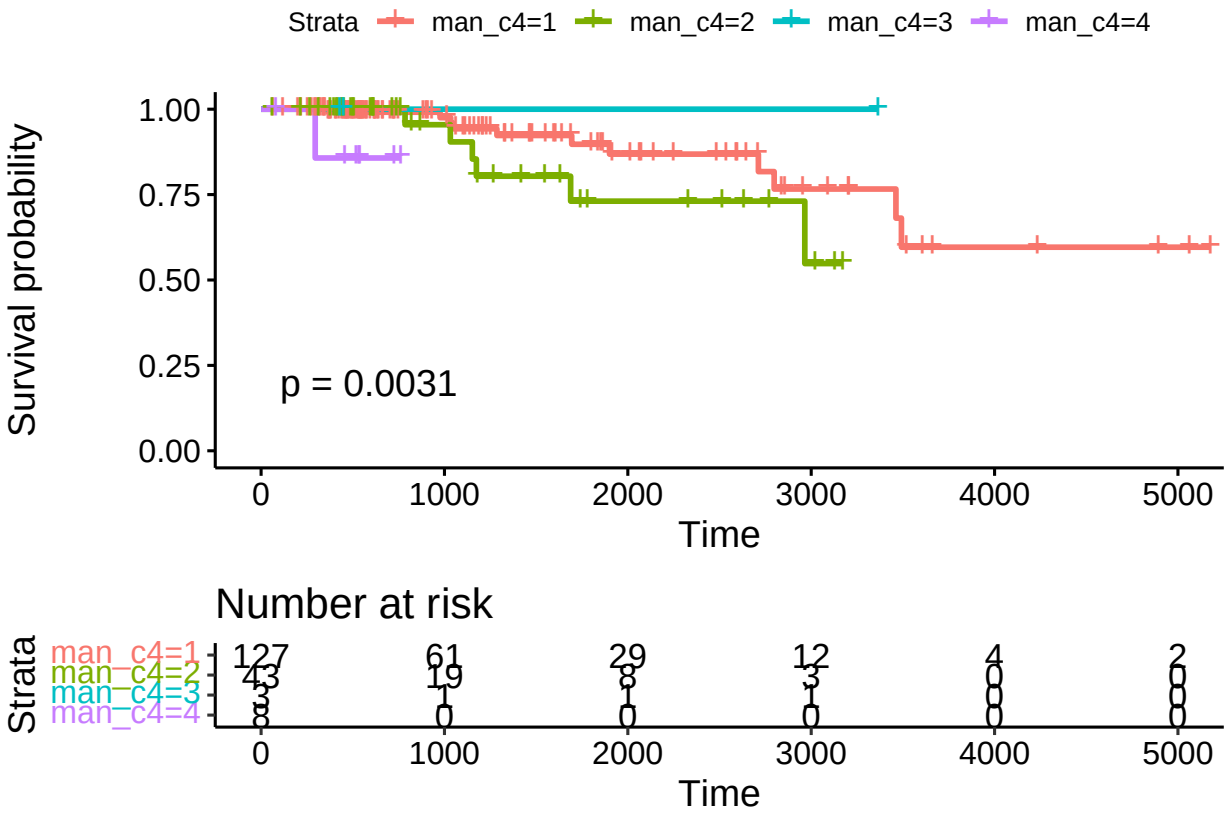
```
ggsurvplot(HC9,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS9,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

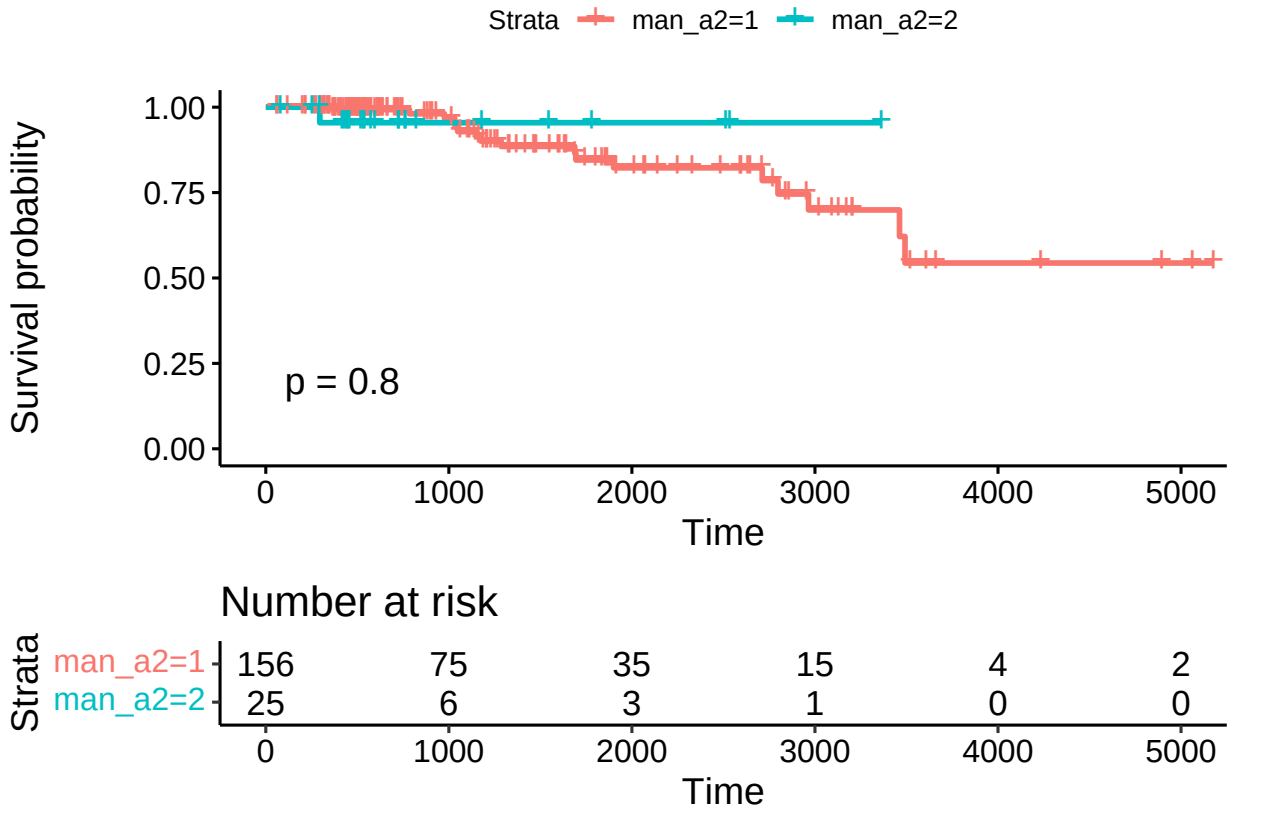



```
ggsurvplot(HW9,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

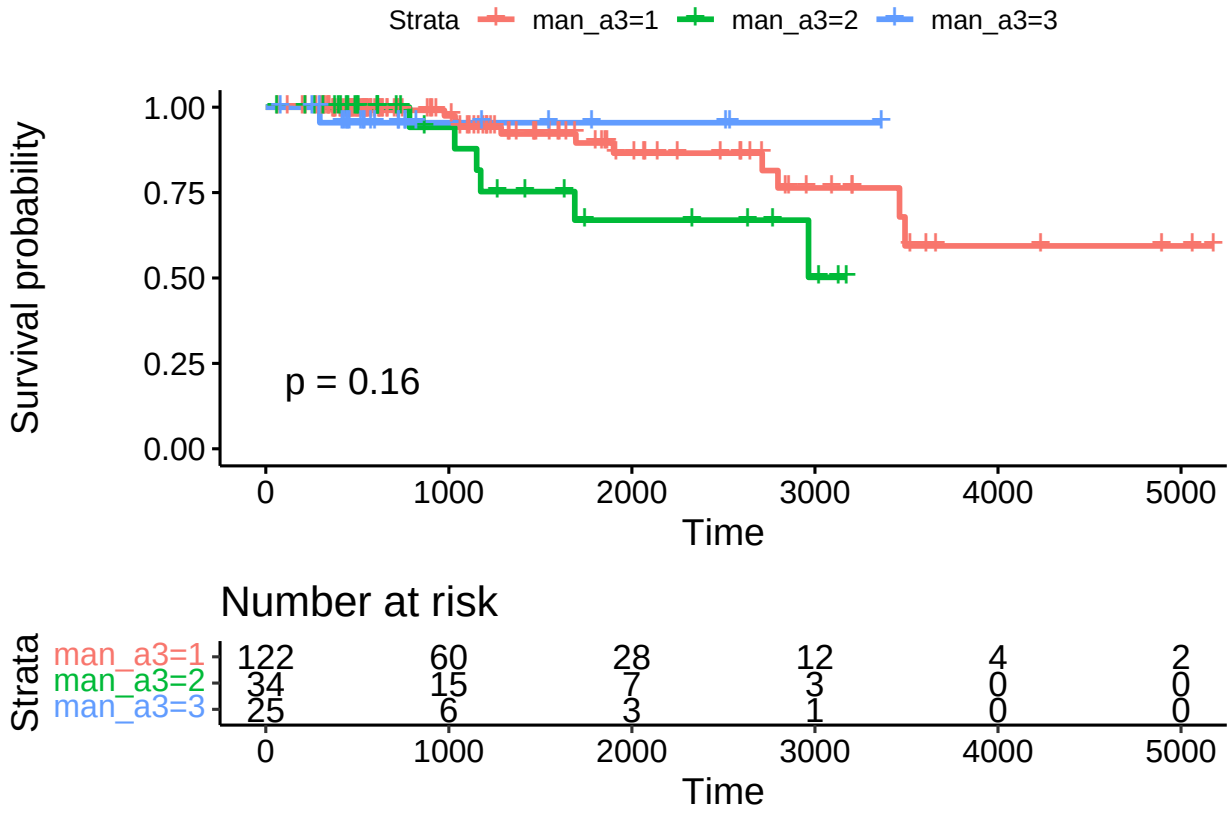


4.2.10. Manhattan distance measurement + Complete Linkage

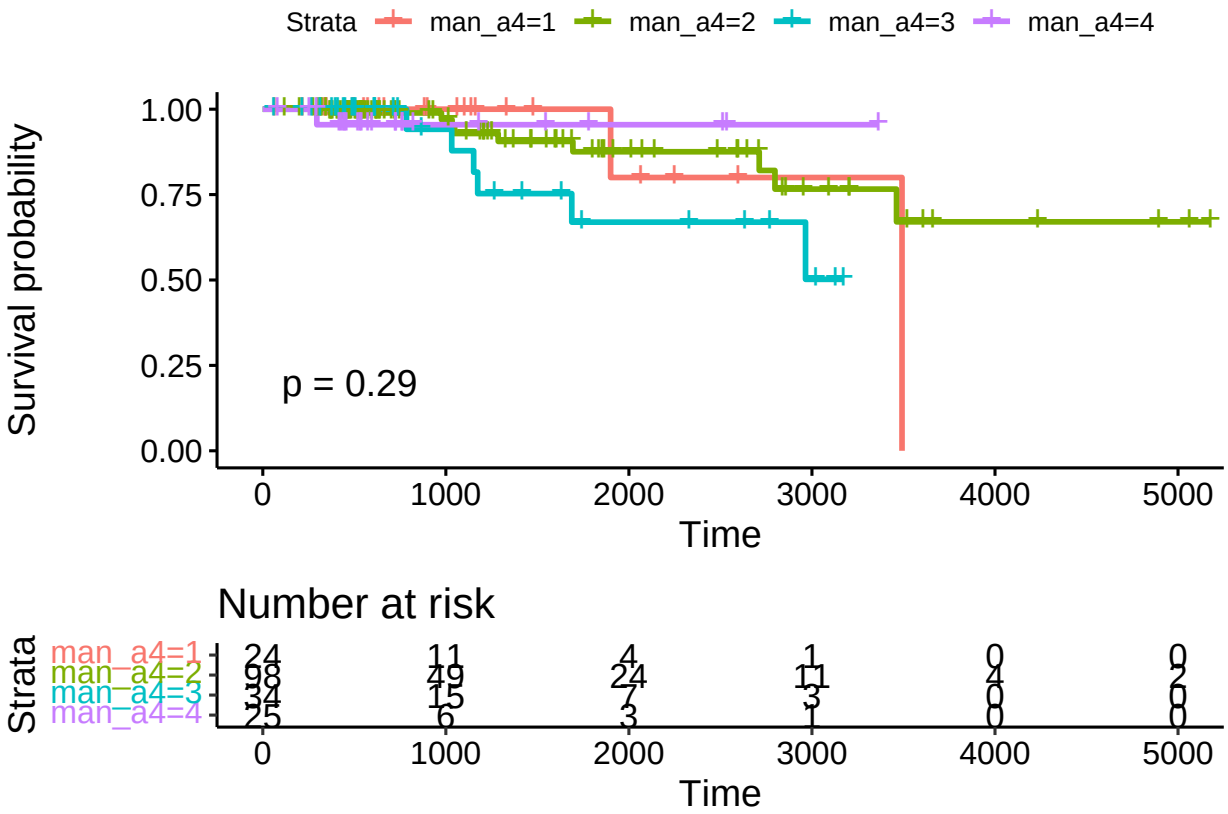
```
ggsurvplot(HC10,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS10,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

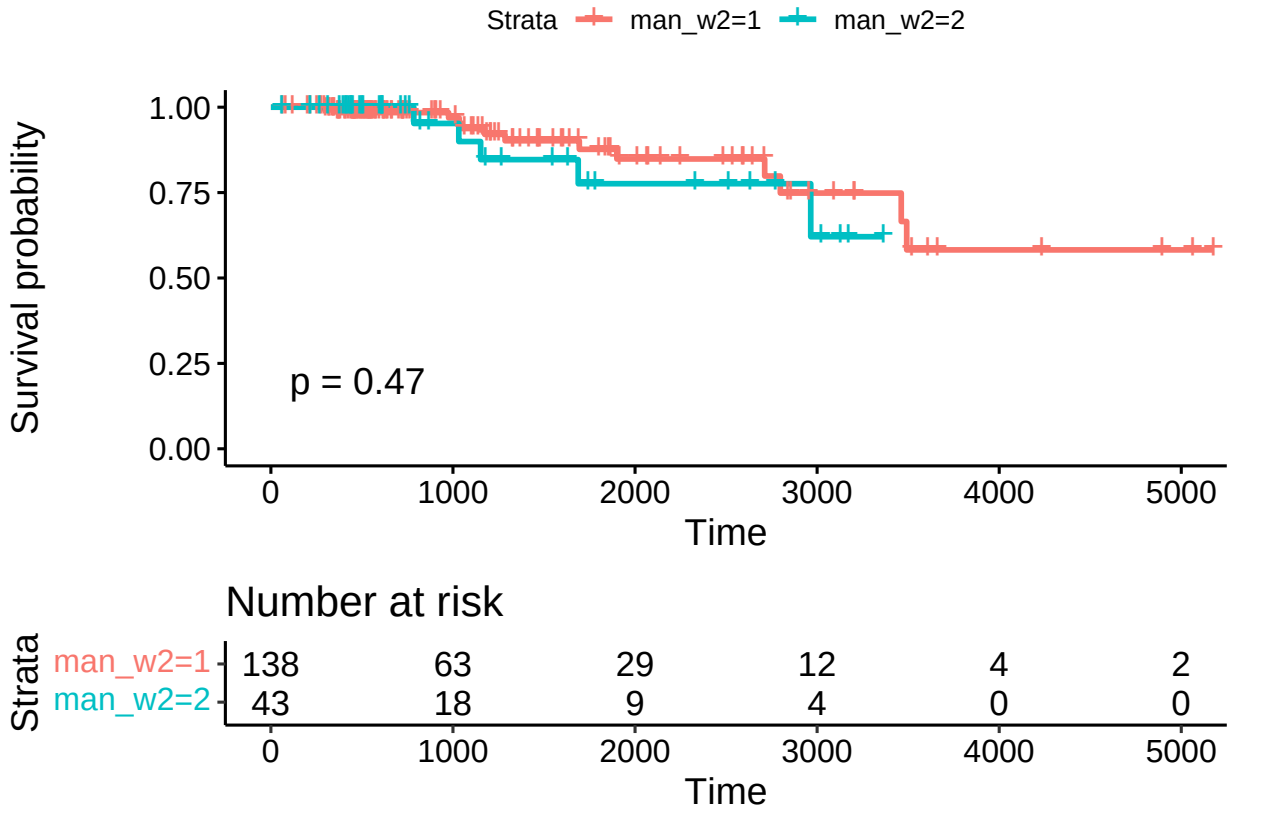



```
ggsurvplot(HW10,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

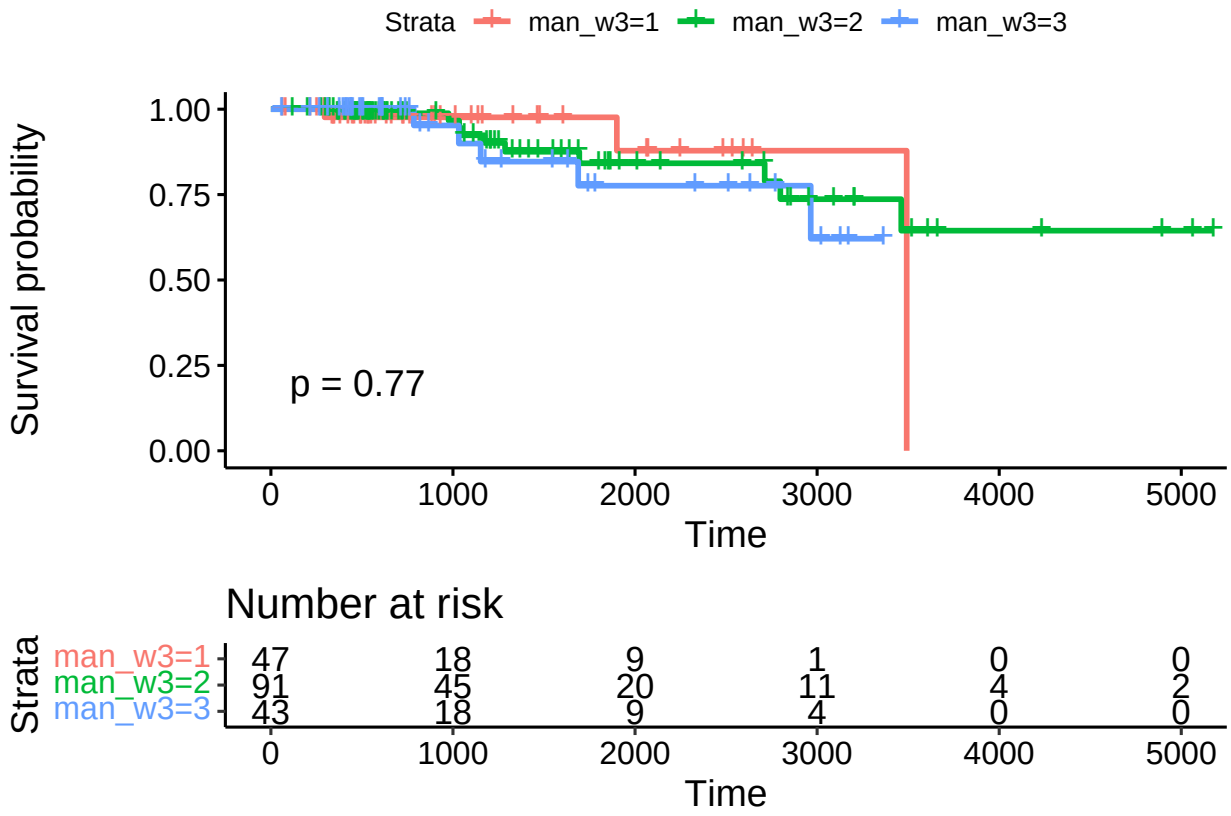


4.2.11. Manhattan distance measurement + Ward Linkage

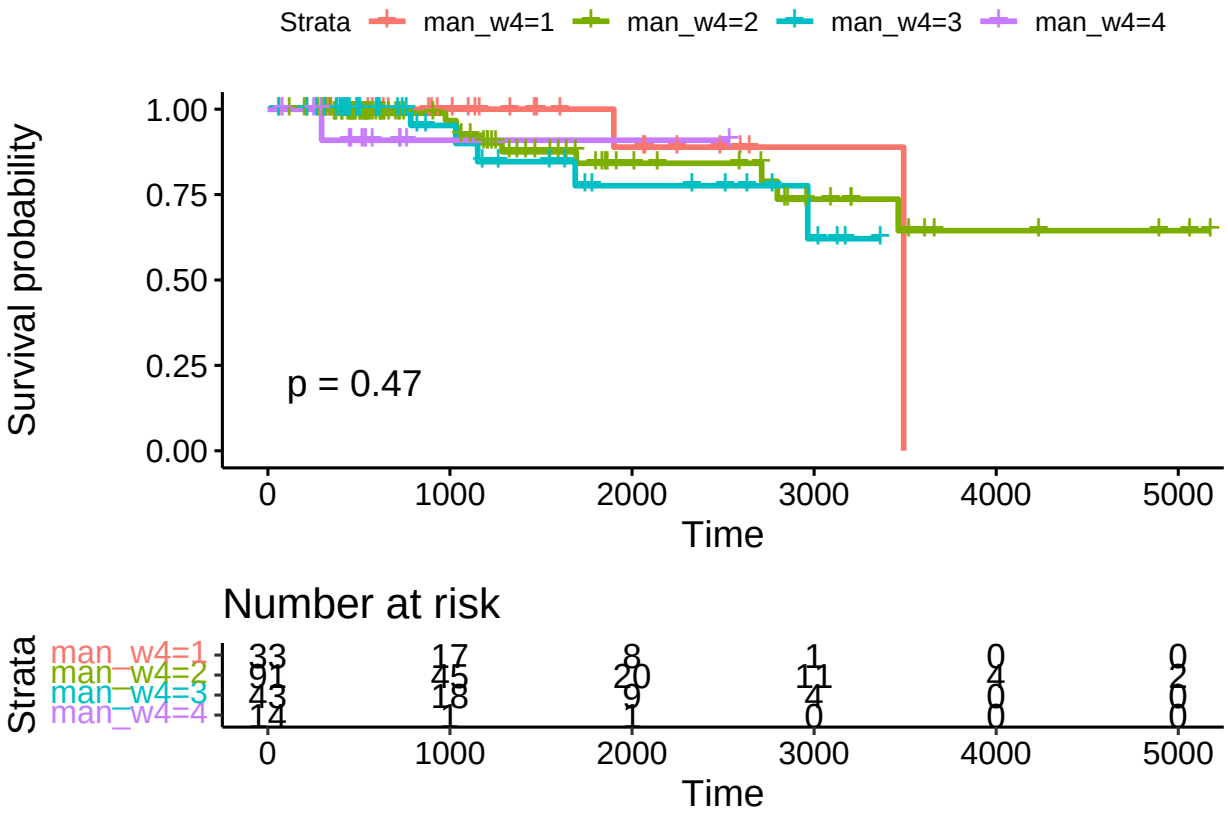
```
ggsurvplot(HC11,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS11,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

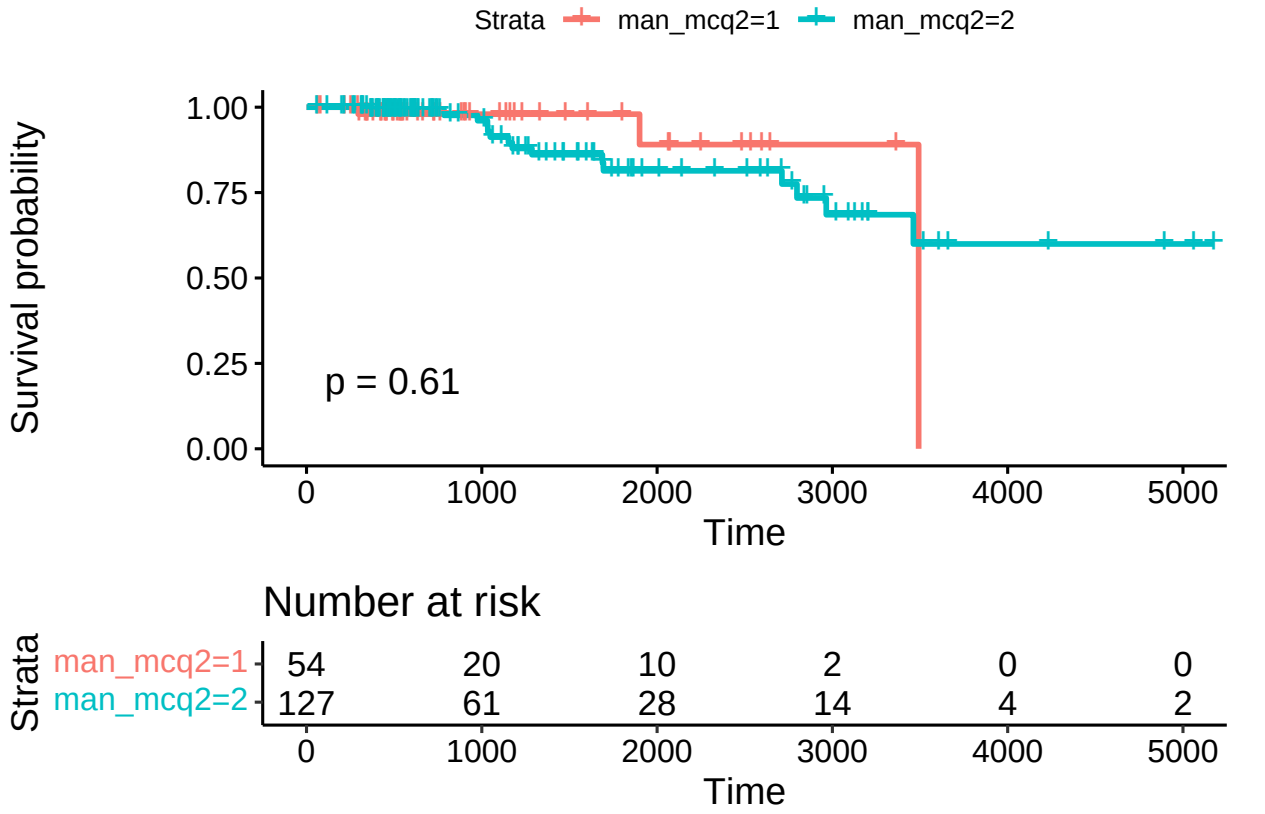



```
ggsurvplot(HW11,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

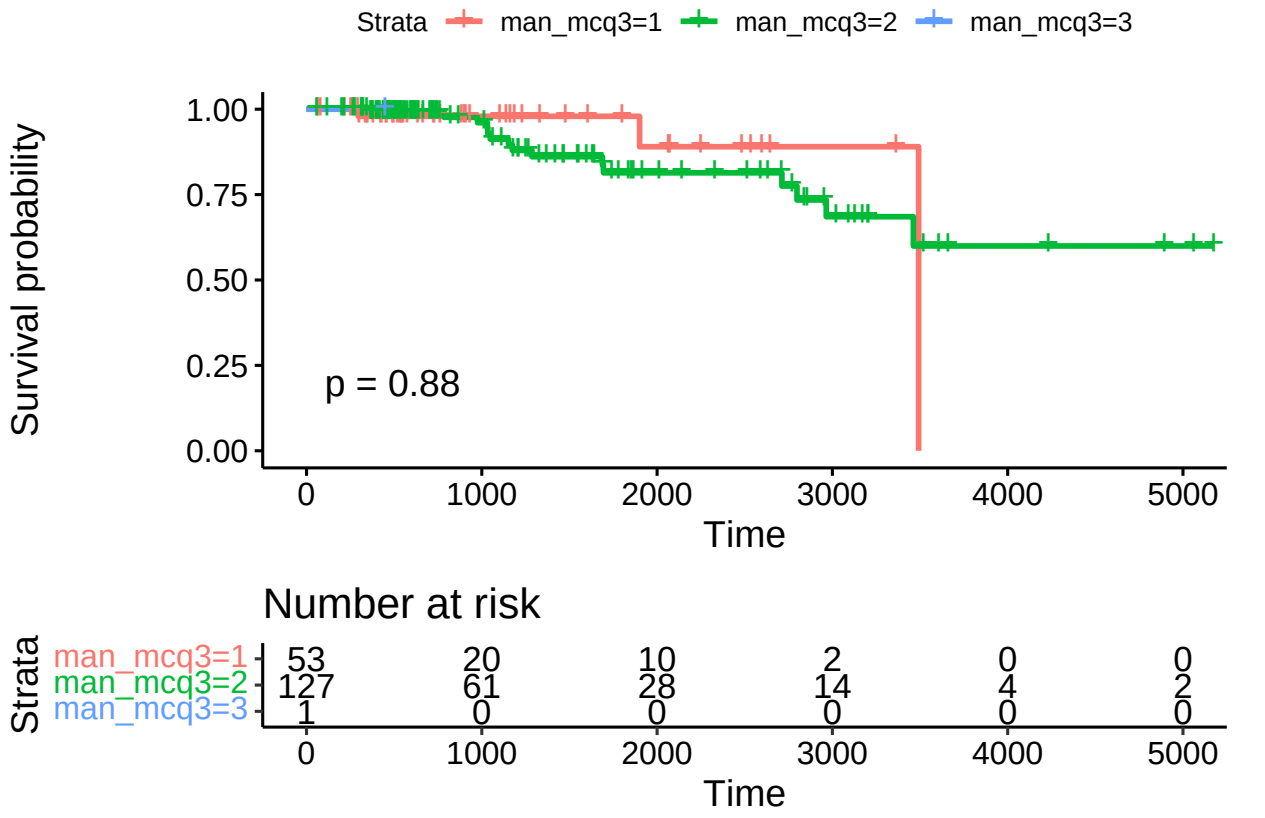


4.2.12. Manhattan distance measurement + Mcquitty Linkage

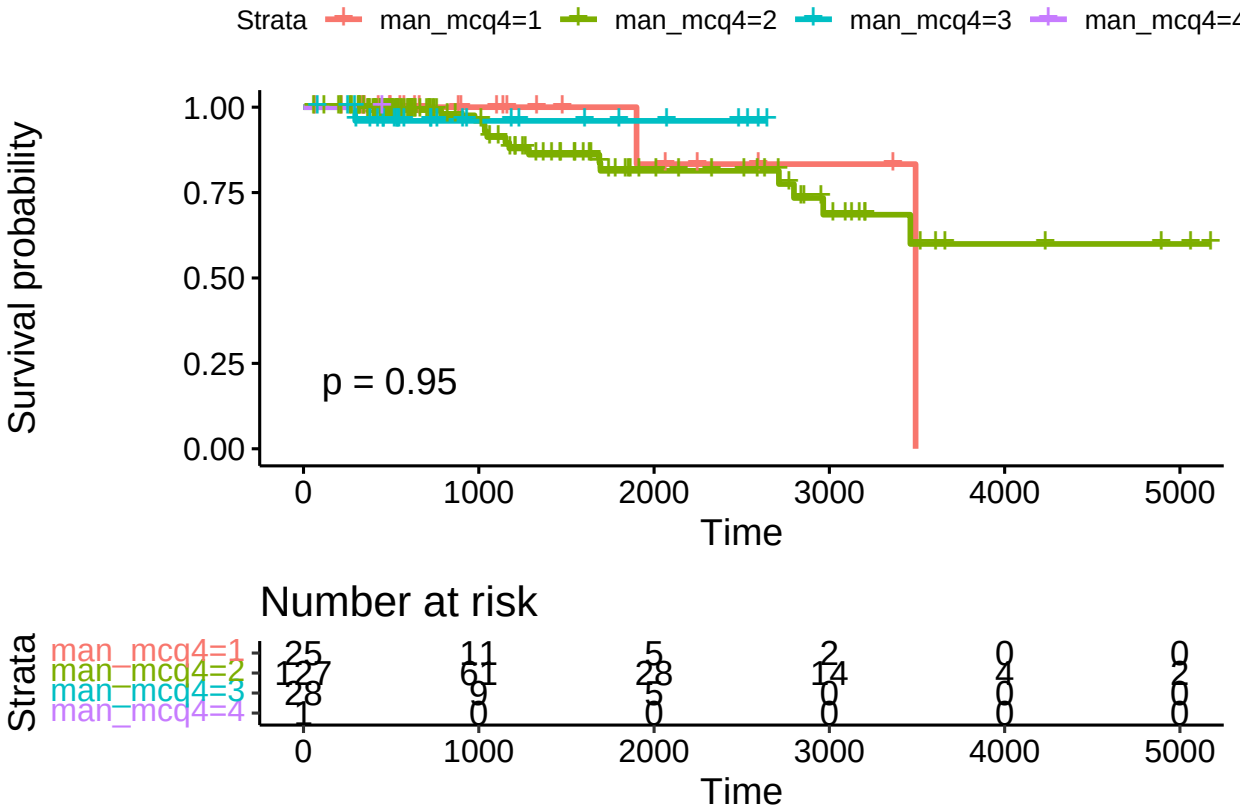
```
ggsurvplot(HC12,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS12,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

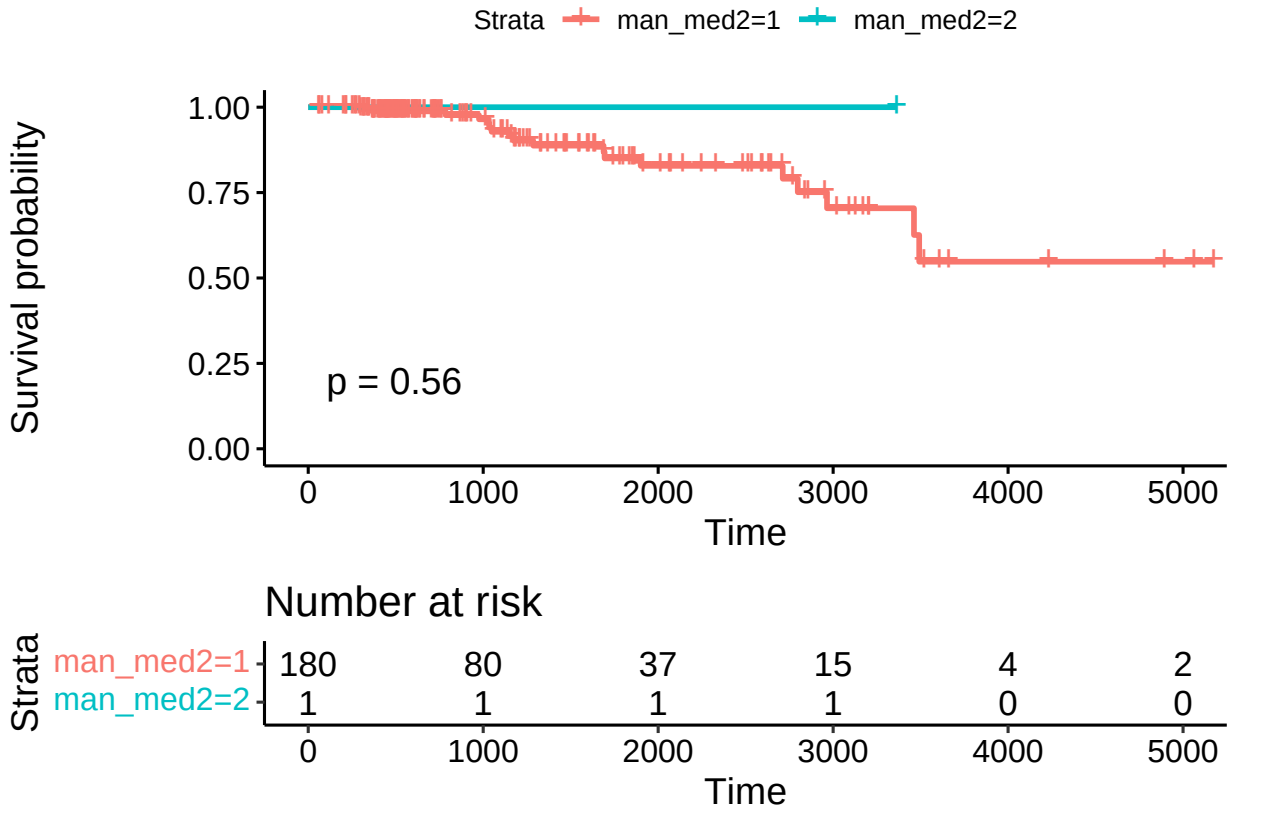



```
ggsurvplot(HW12,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

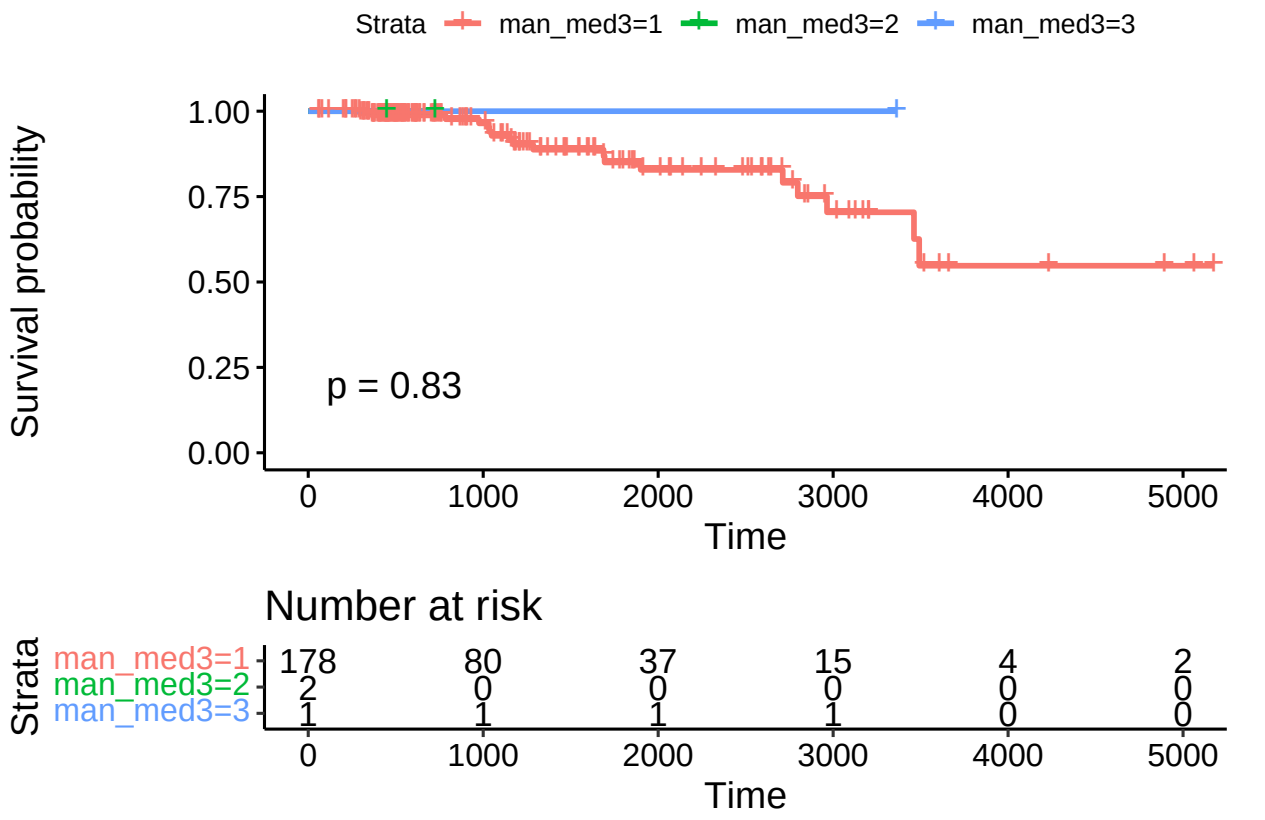


4.2.13. Manhattan distance measurement + Median Linkage

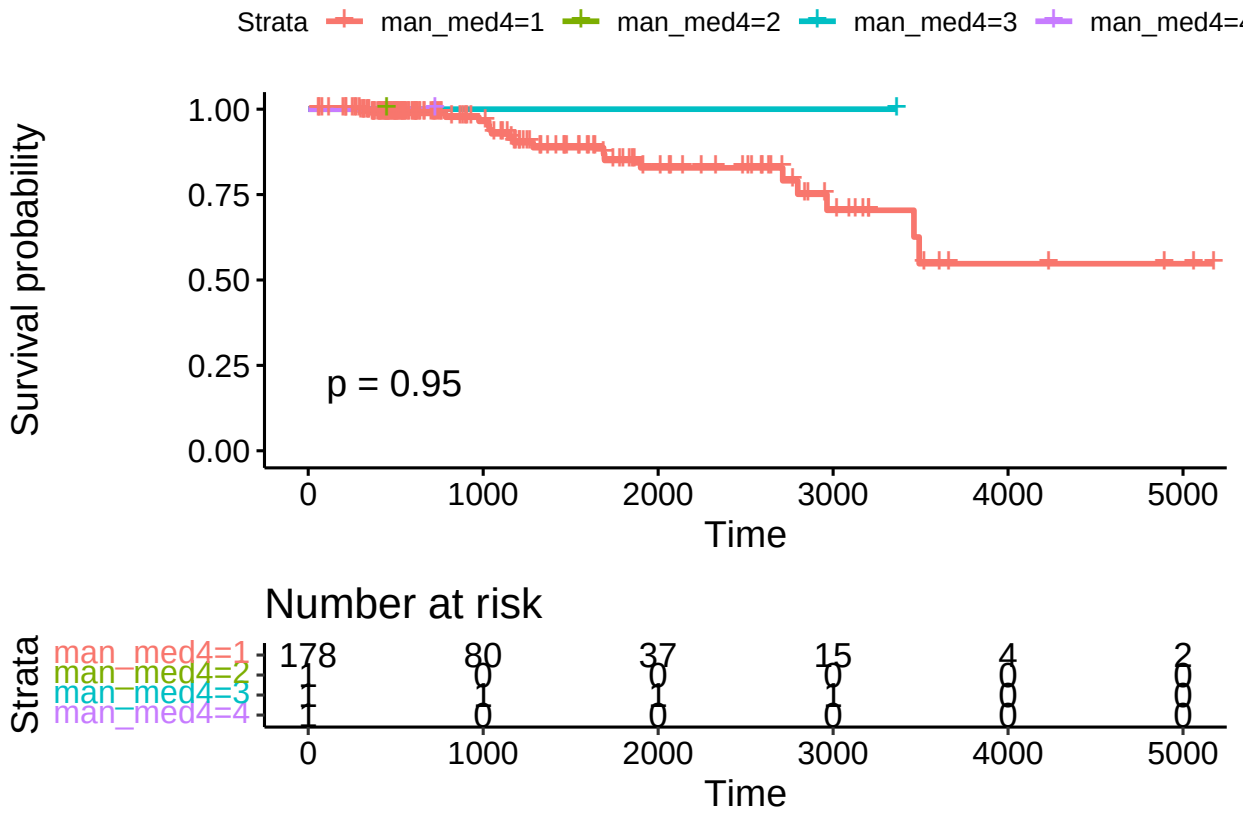
```
ggsurvplot(HC13,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS13,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

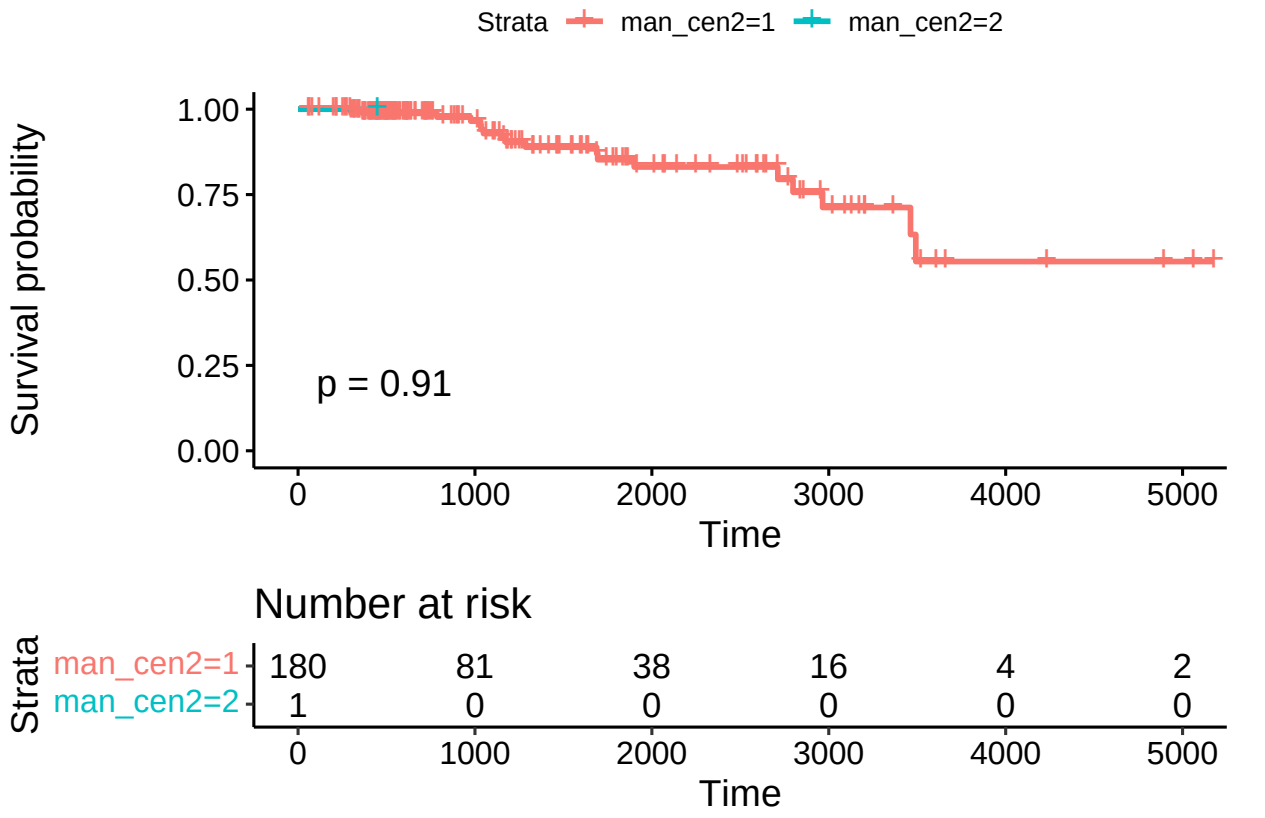



```
ggsurvplot(HW13,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

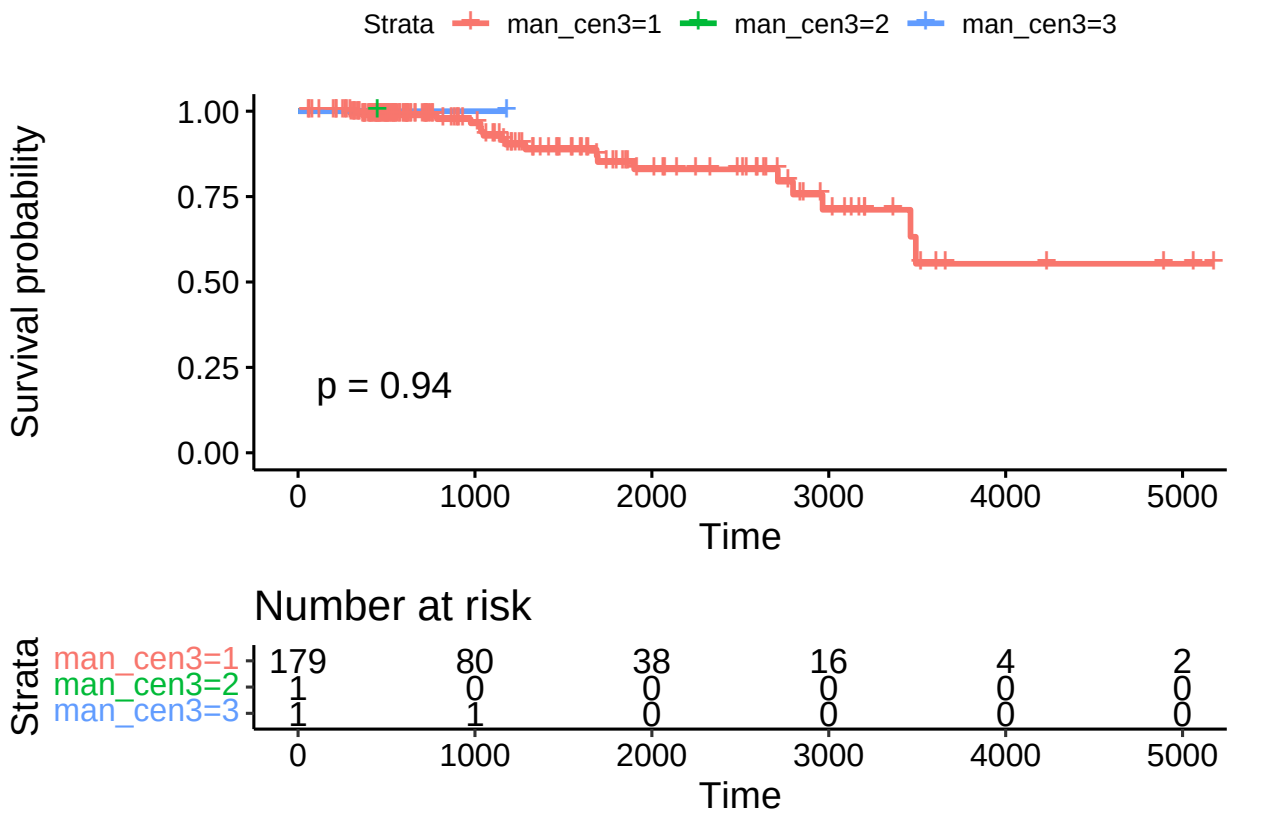


4.2.14. Manhattan distance measurement + Centroid Linkage

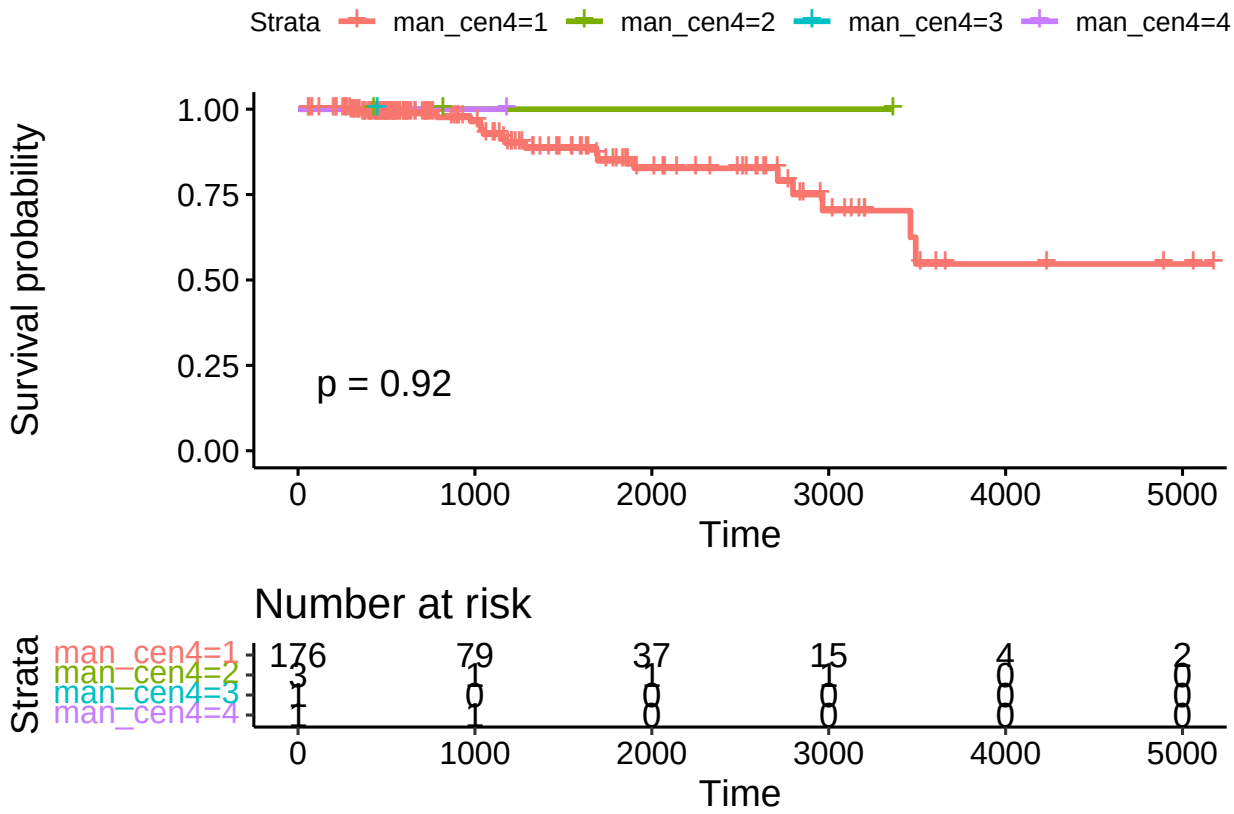
```
ggsurvplot(HC14,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS14,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

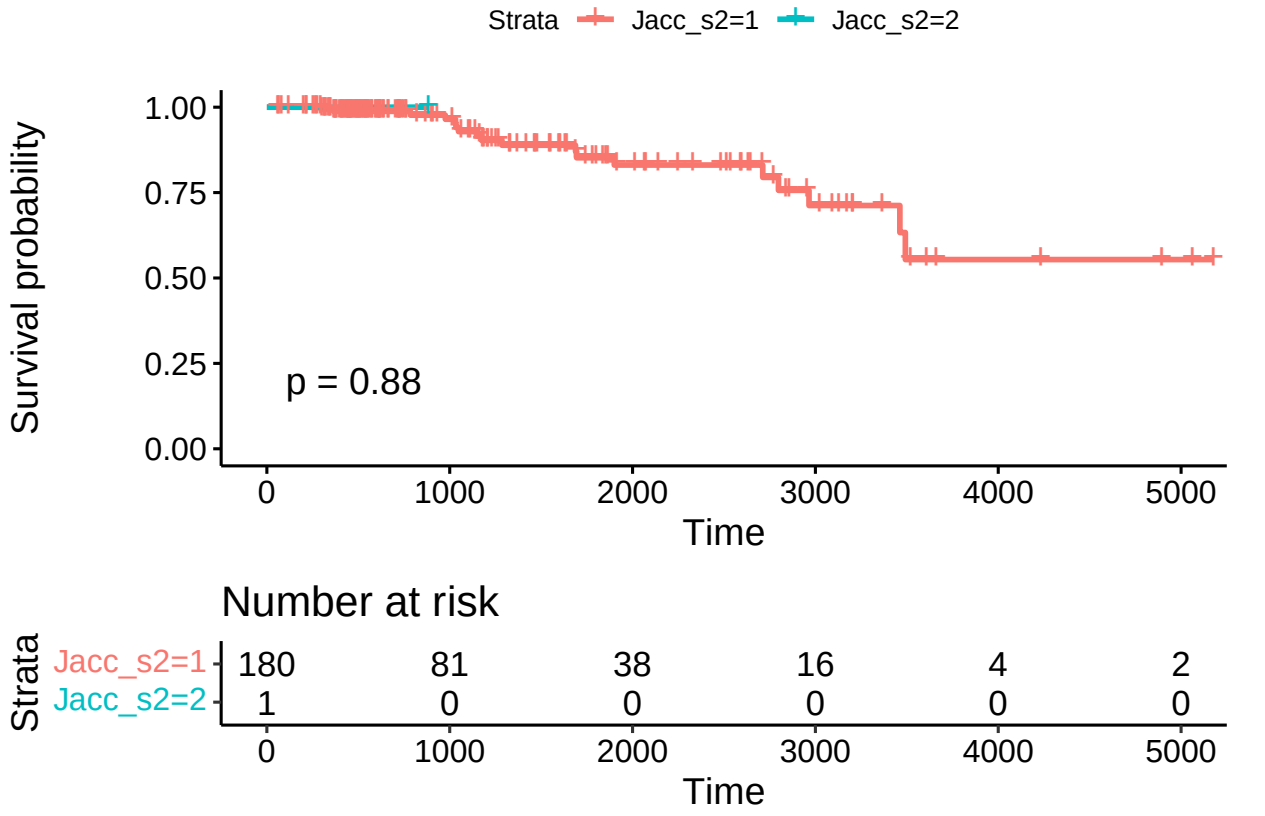



```
ggsurvplot(HW14,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

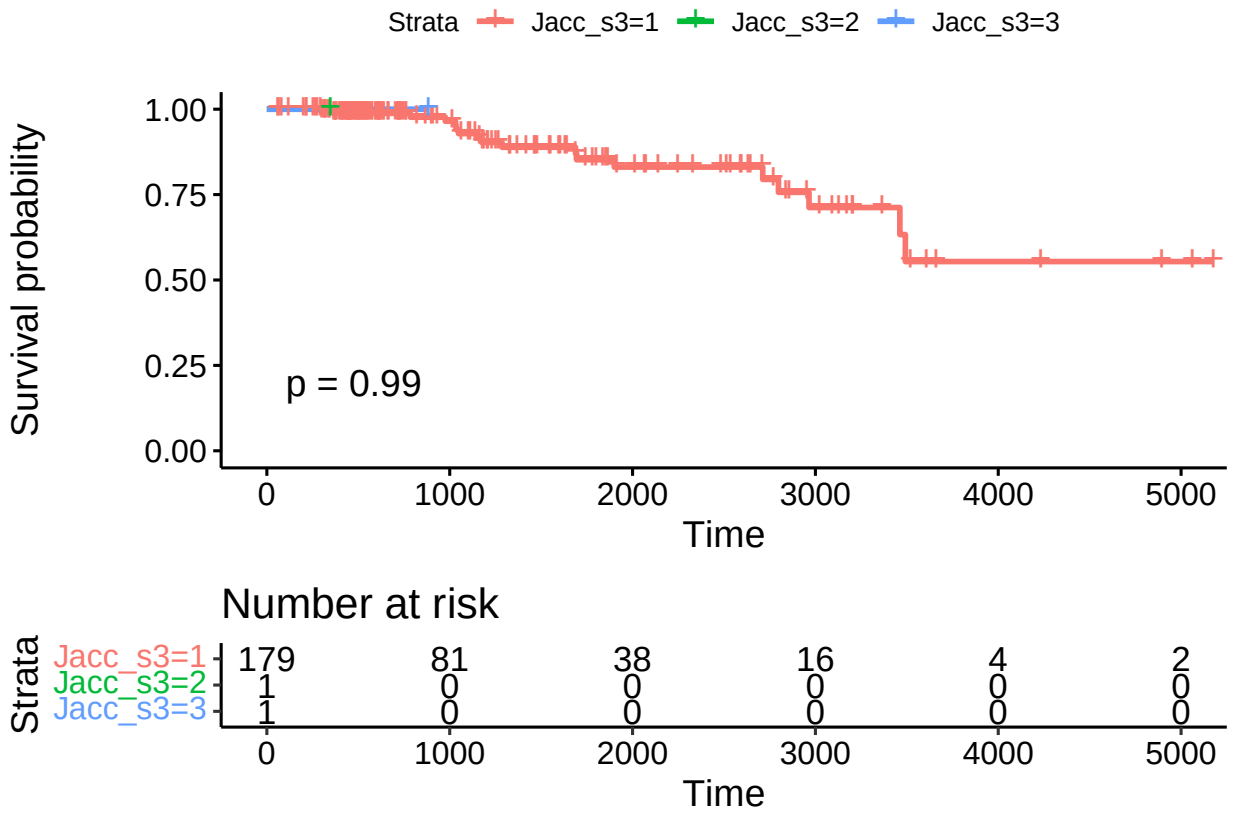


4.2.15. Jaccard distance measurement + Single Linkage

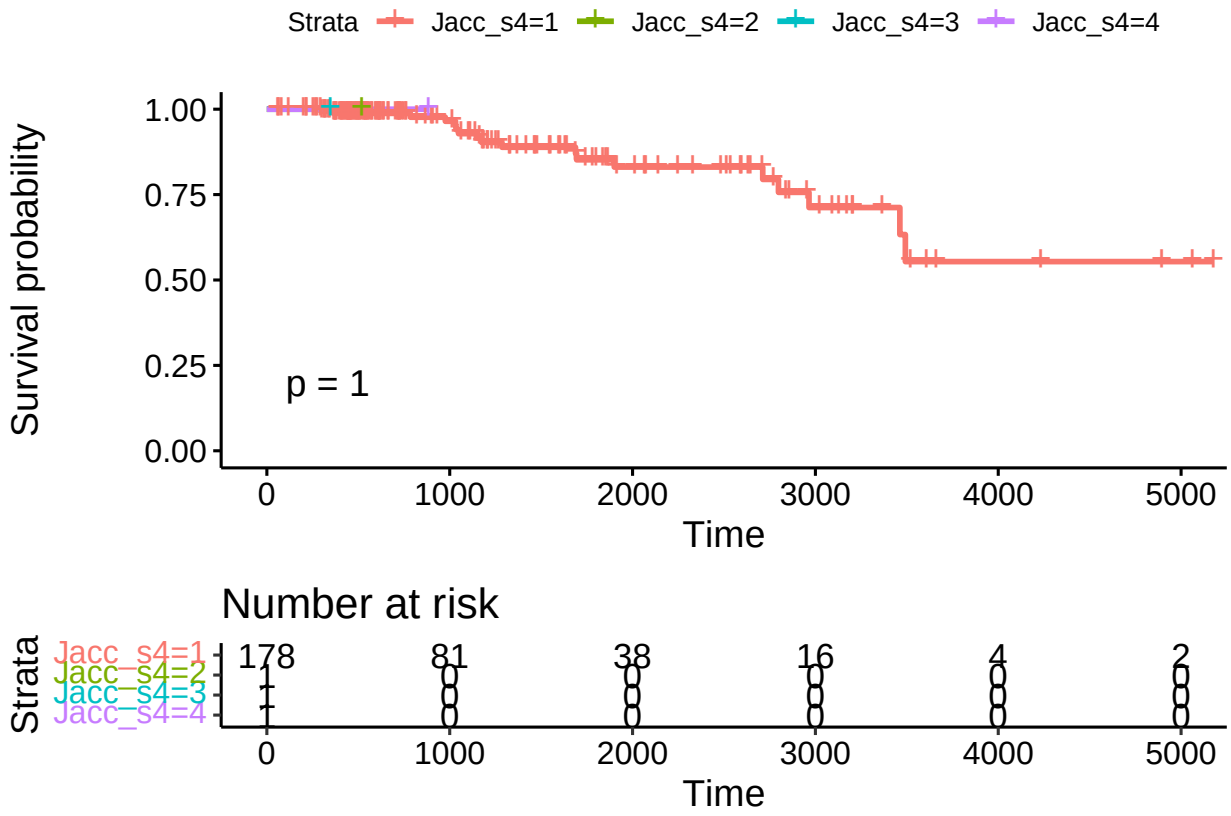
```
ggsurvplot(HC15,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS15,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

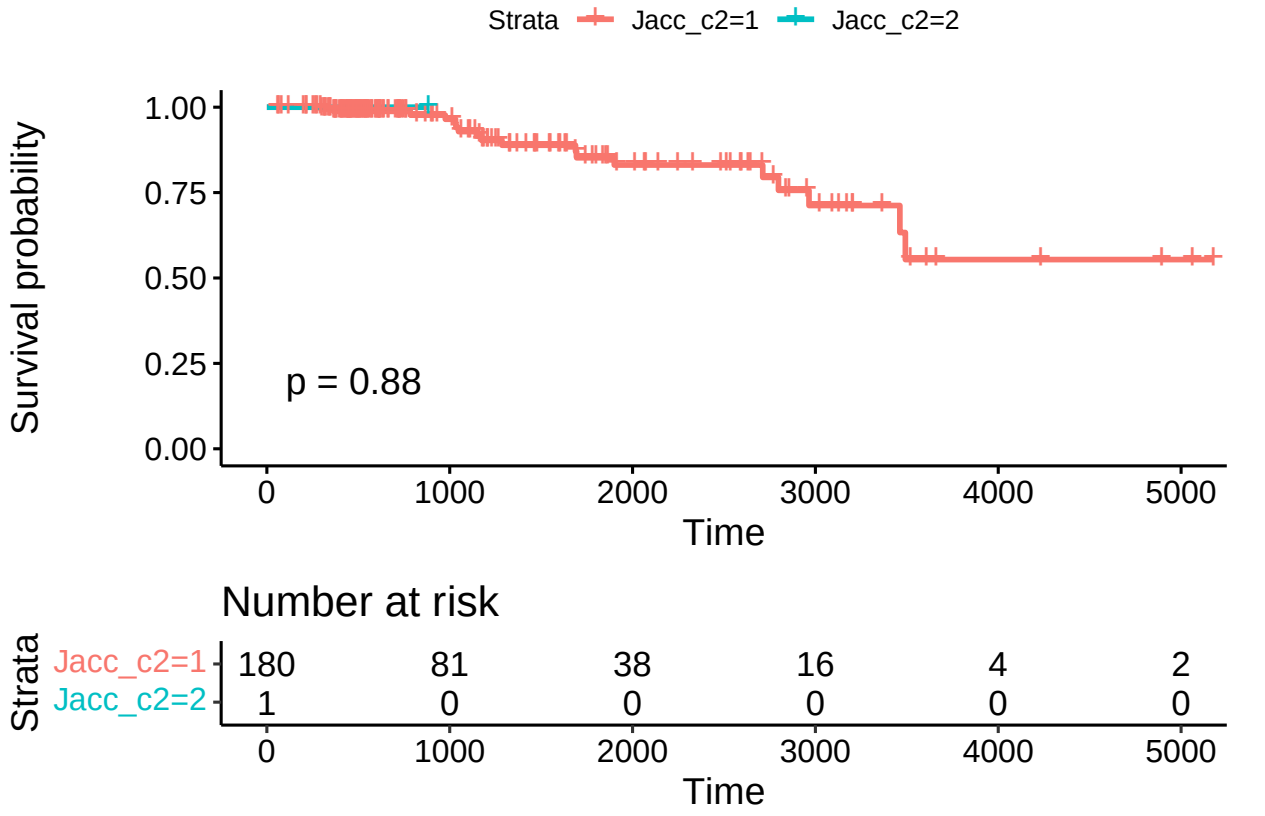



```
ggsurvplot(HW15,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

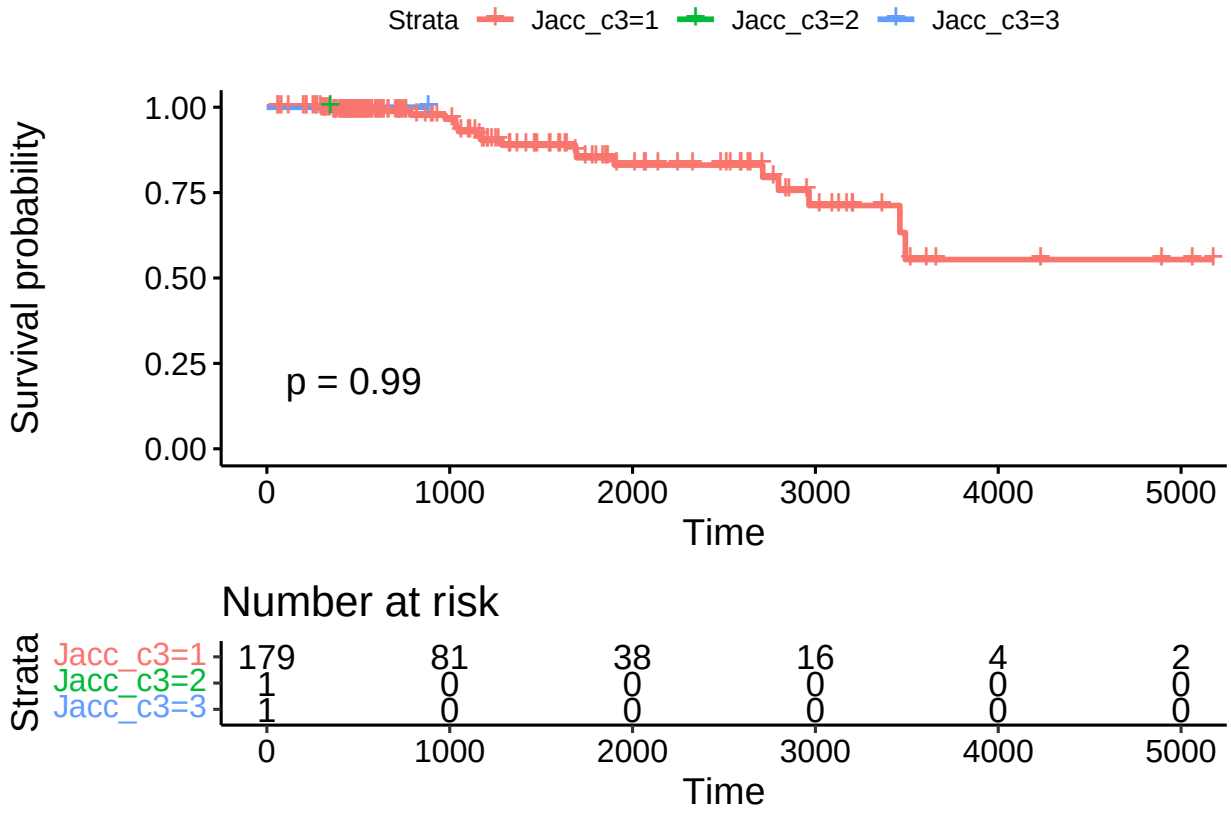


4.2.16. Jaccard distance measurement + Average Linkage

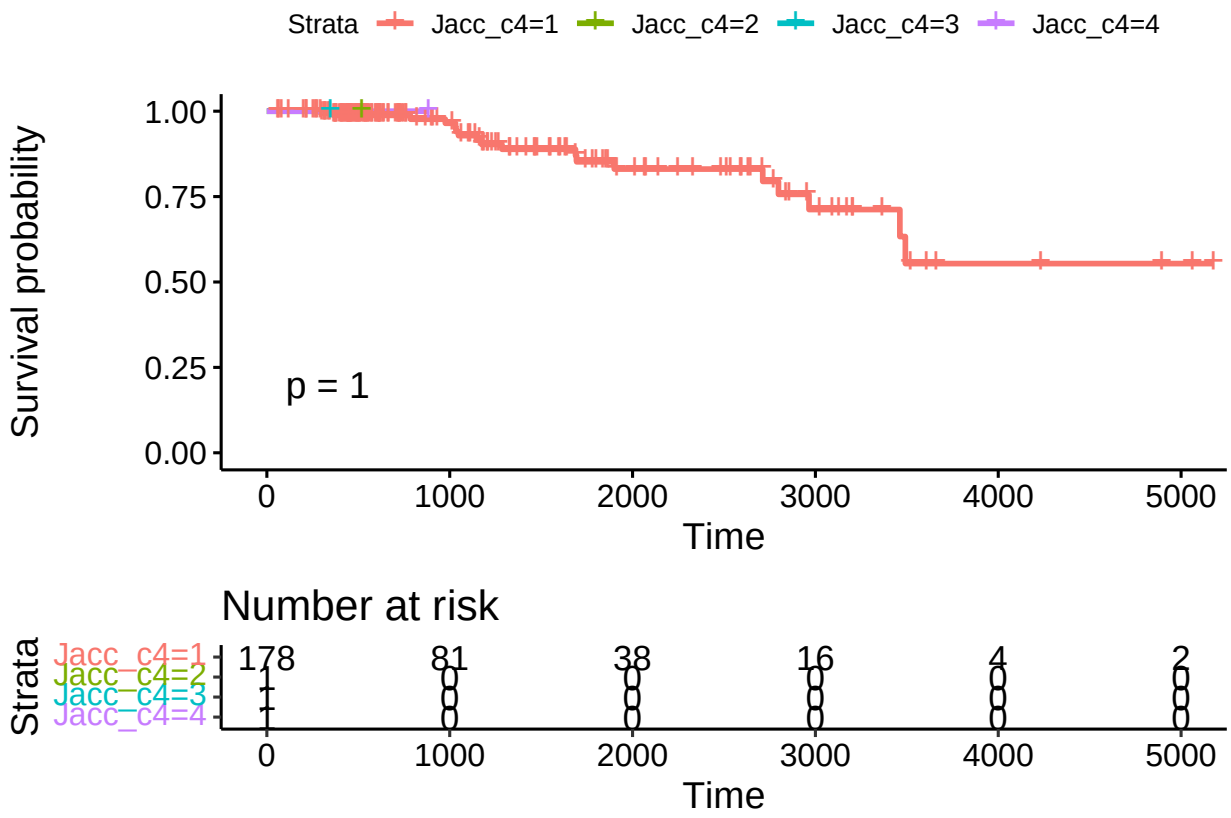
```
ggsurvplot(HC16,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS16,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

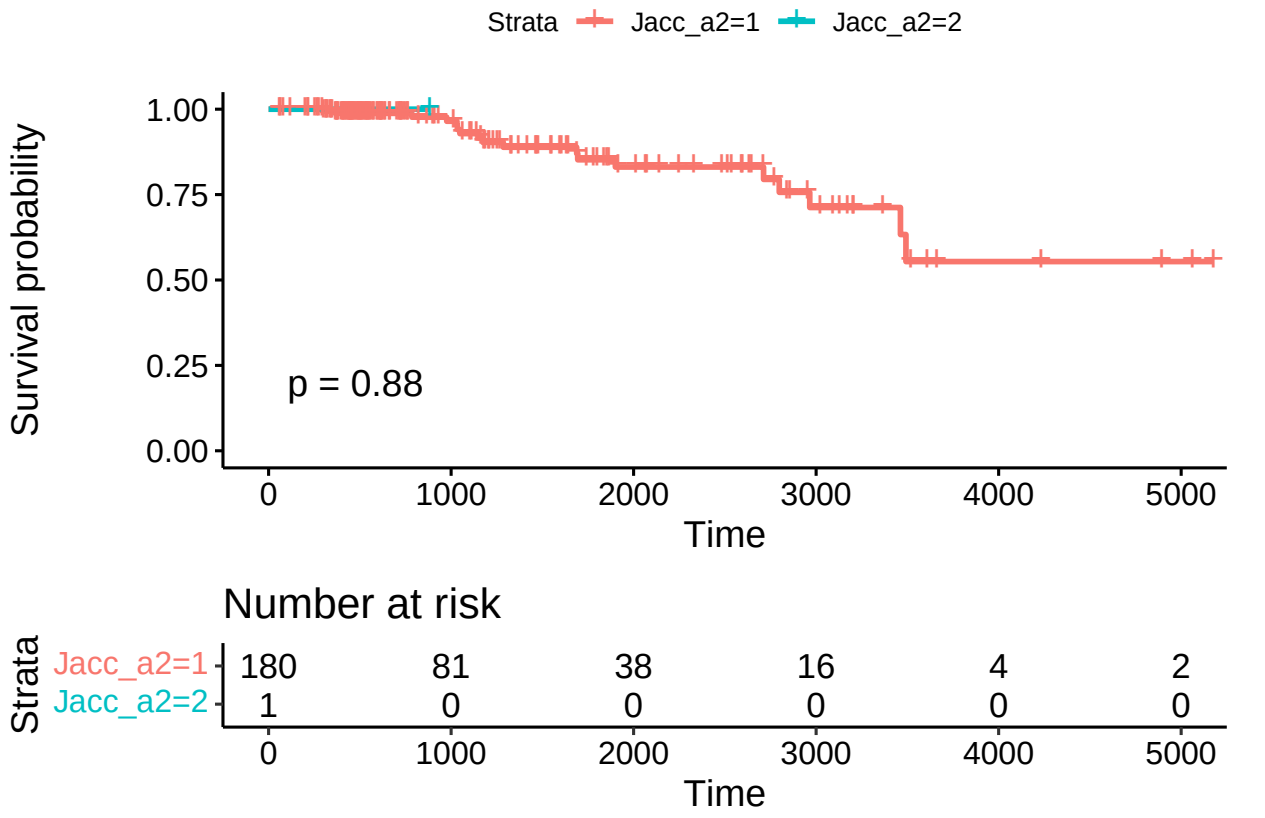



```
ggsurvplot(HW16,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

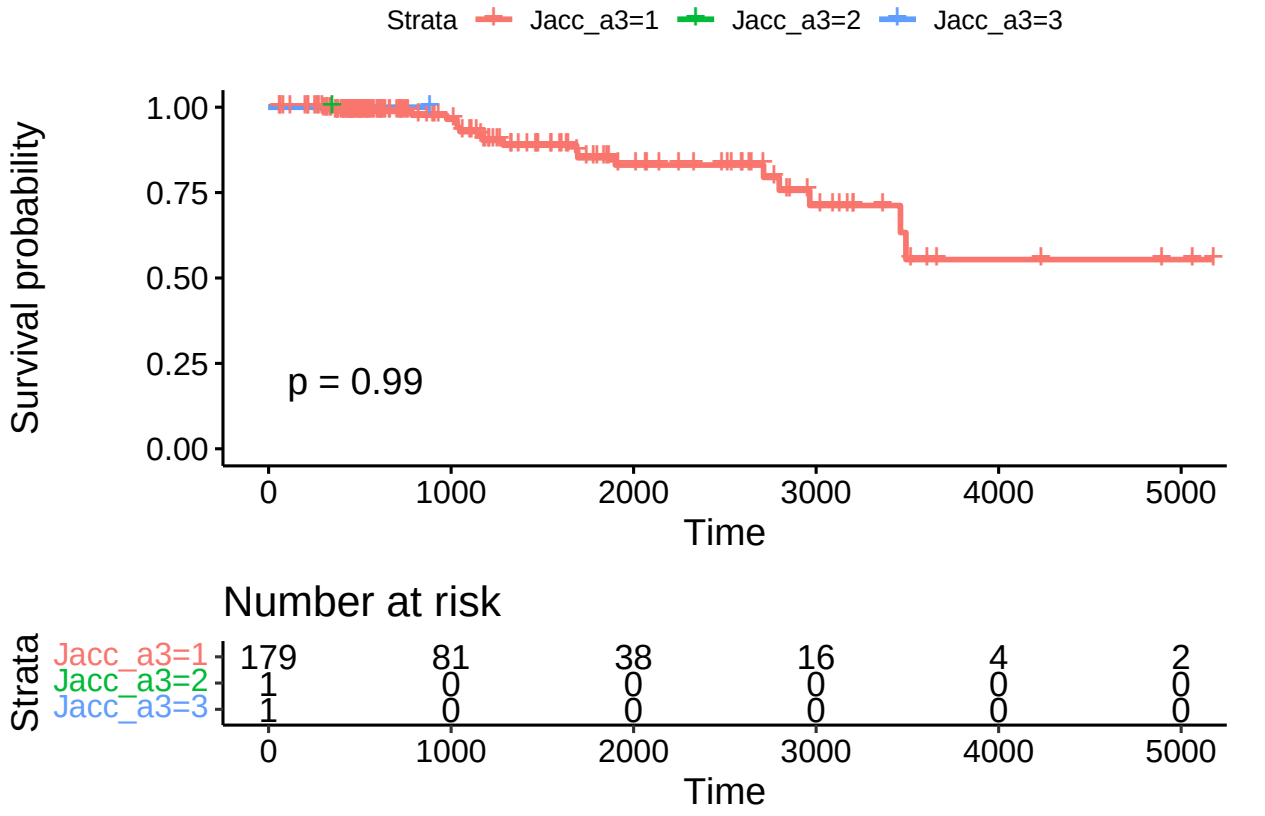


4.2.17. Jaccard distance measurement + Complete Linkage

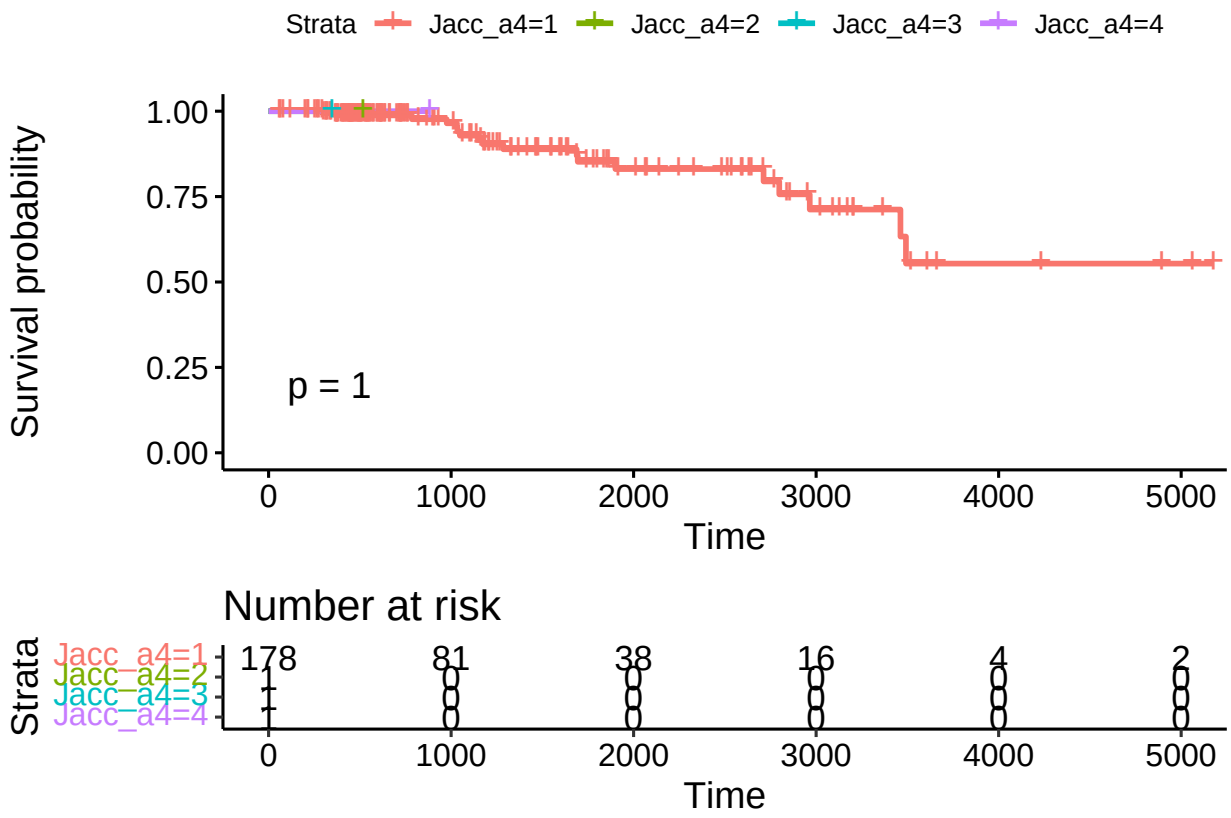
```
ggsurvplot(HC17,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS17,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

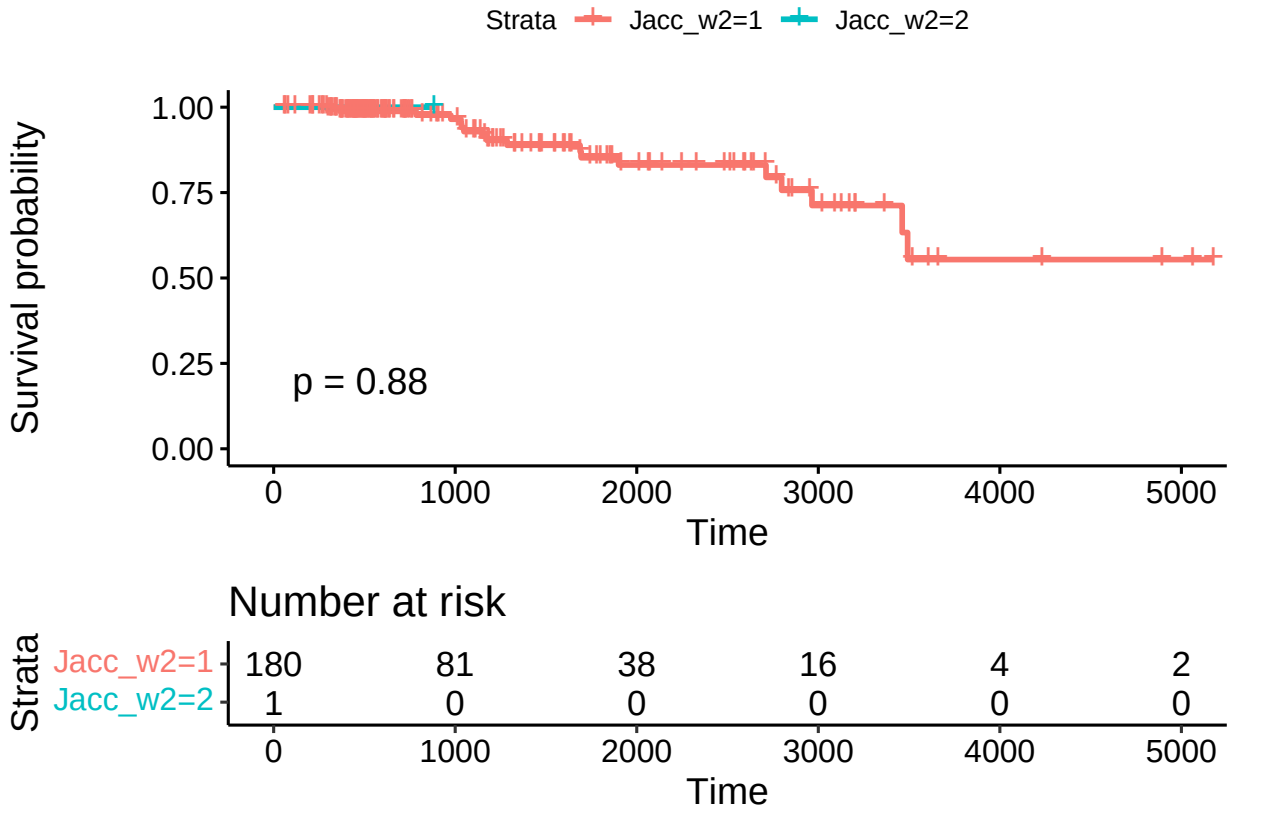



```
ggsurvplot(HW17,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

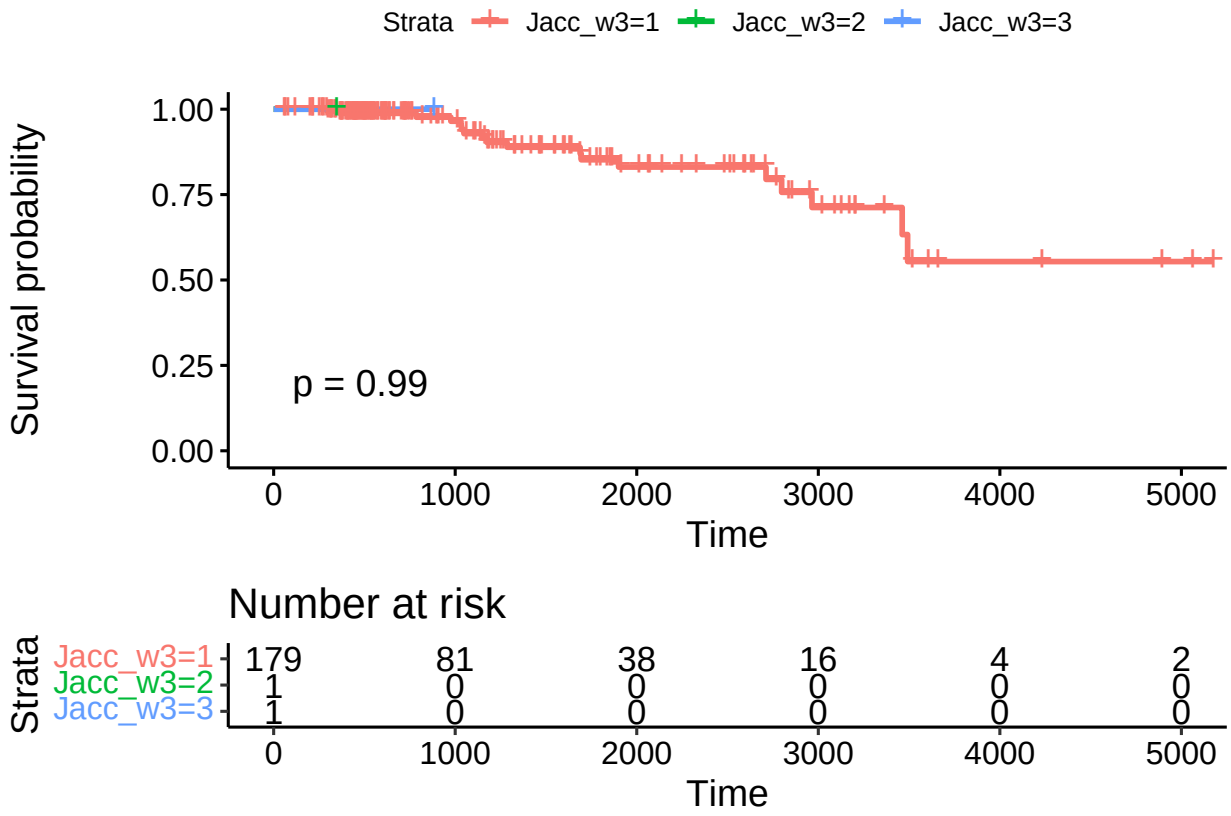


4.2.18. Jaccard distance measurement + Ward Linkage

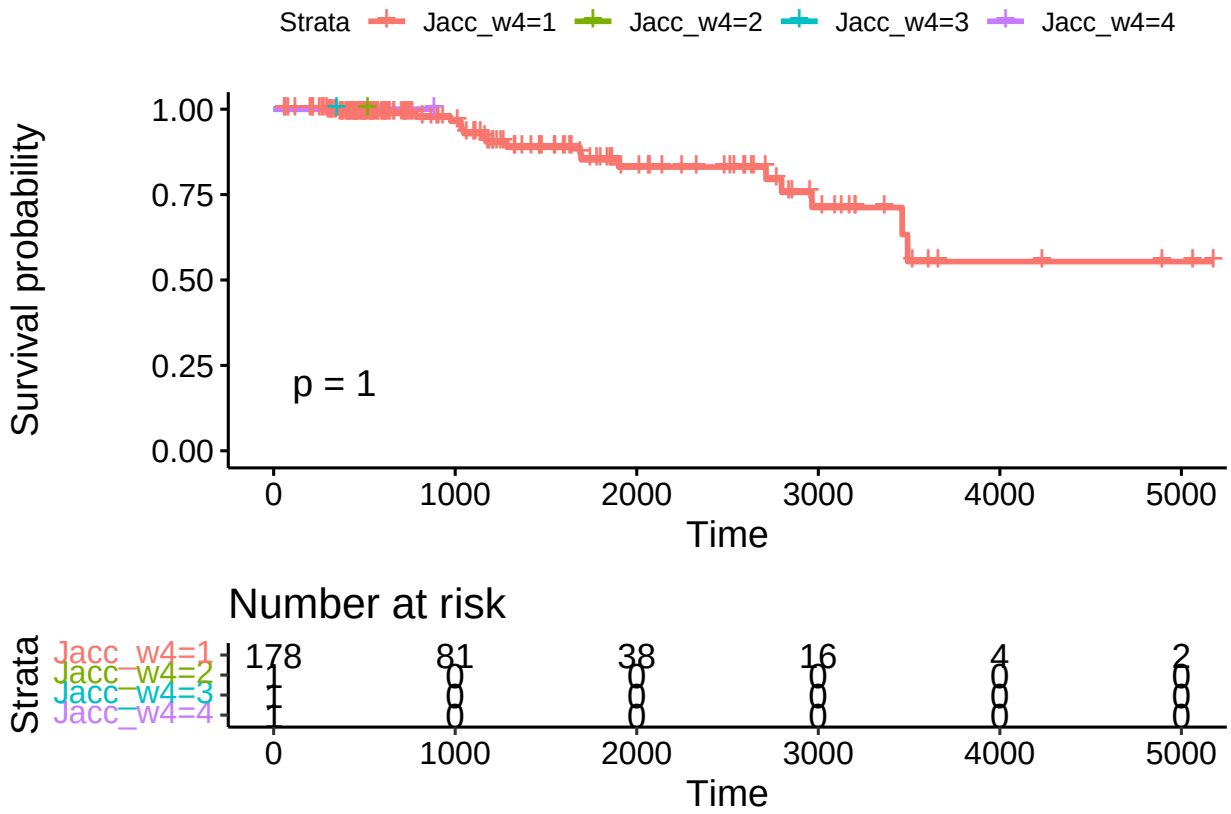
```
ggsurvplot(HC18,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS18,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

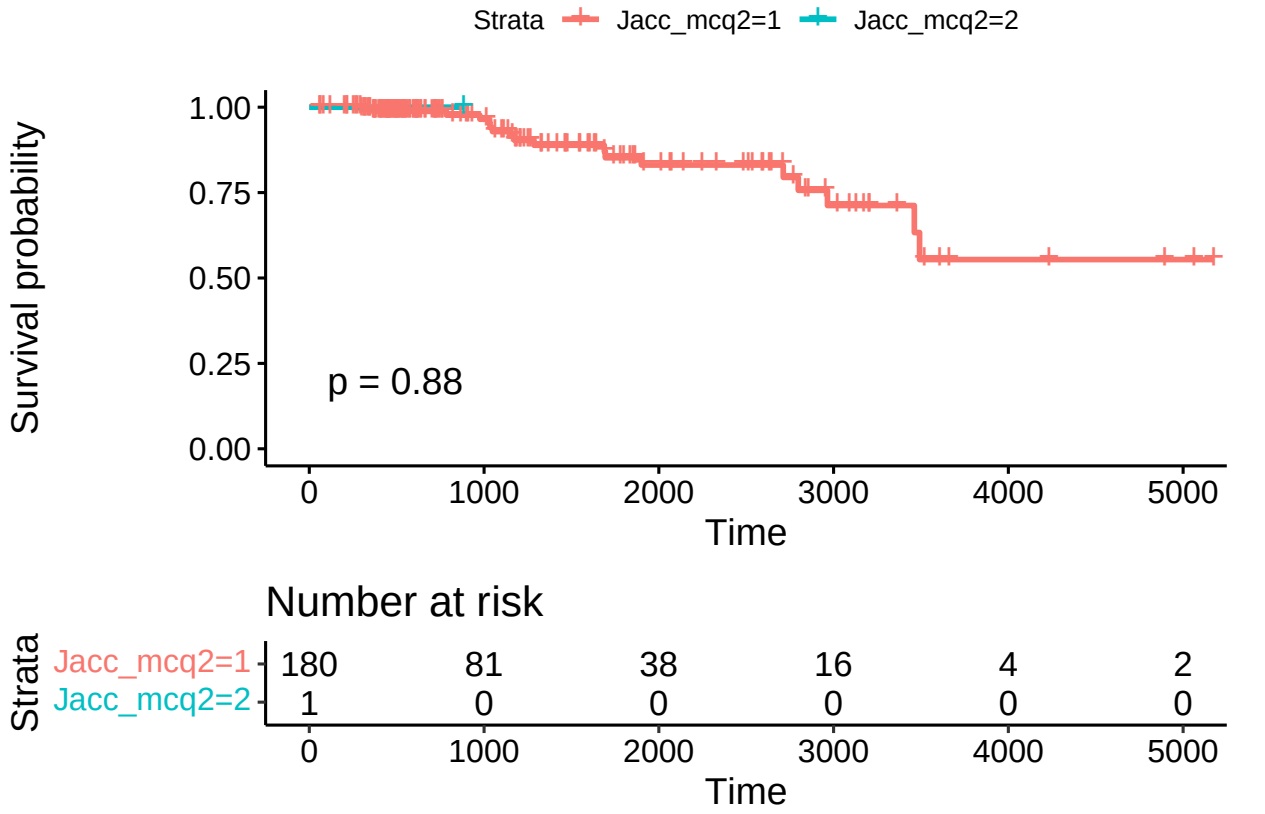



```
ggsurvplot(HW18,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

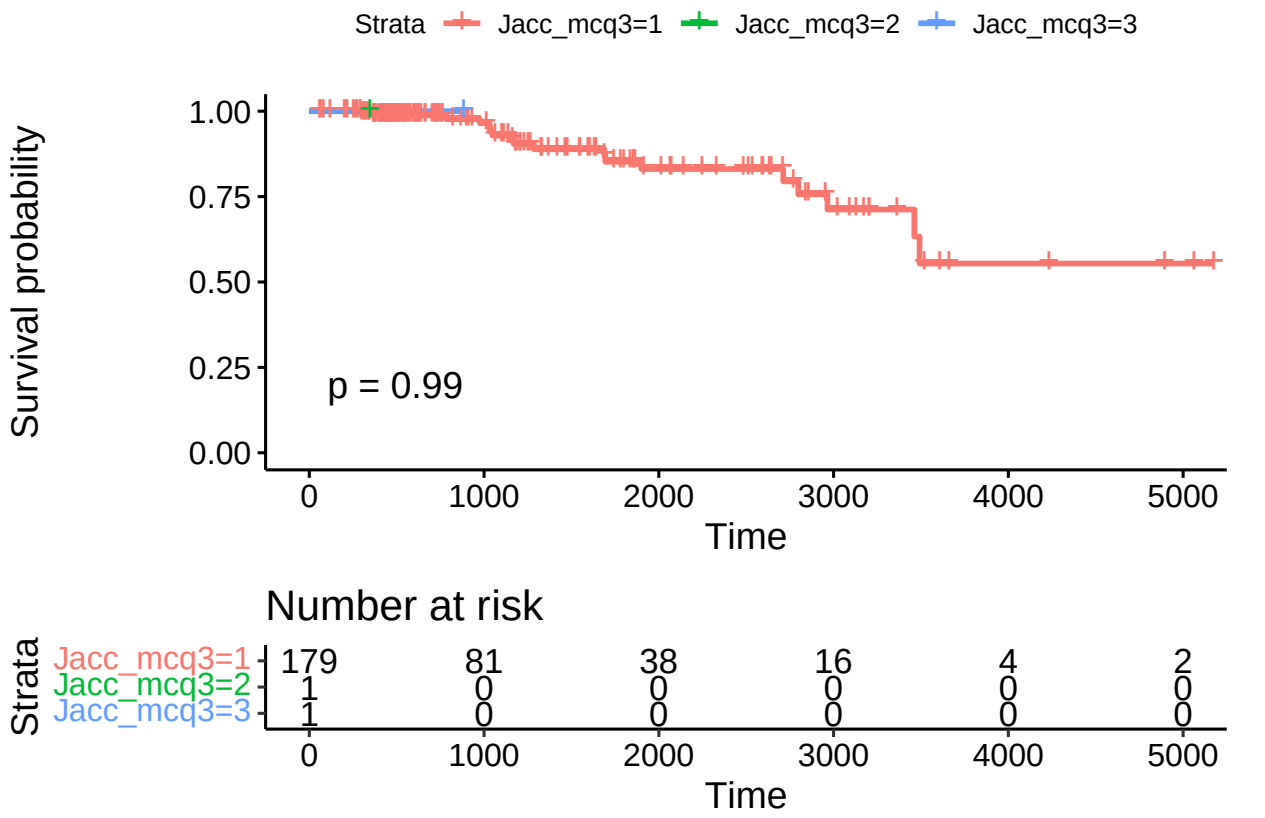


4.2.19. Jaccard distance measurement + Mcquitty Linkage

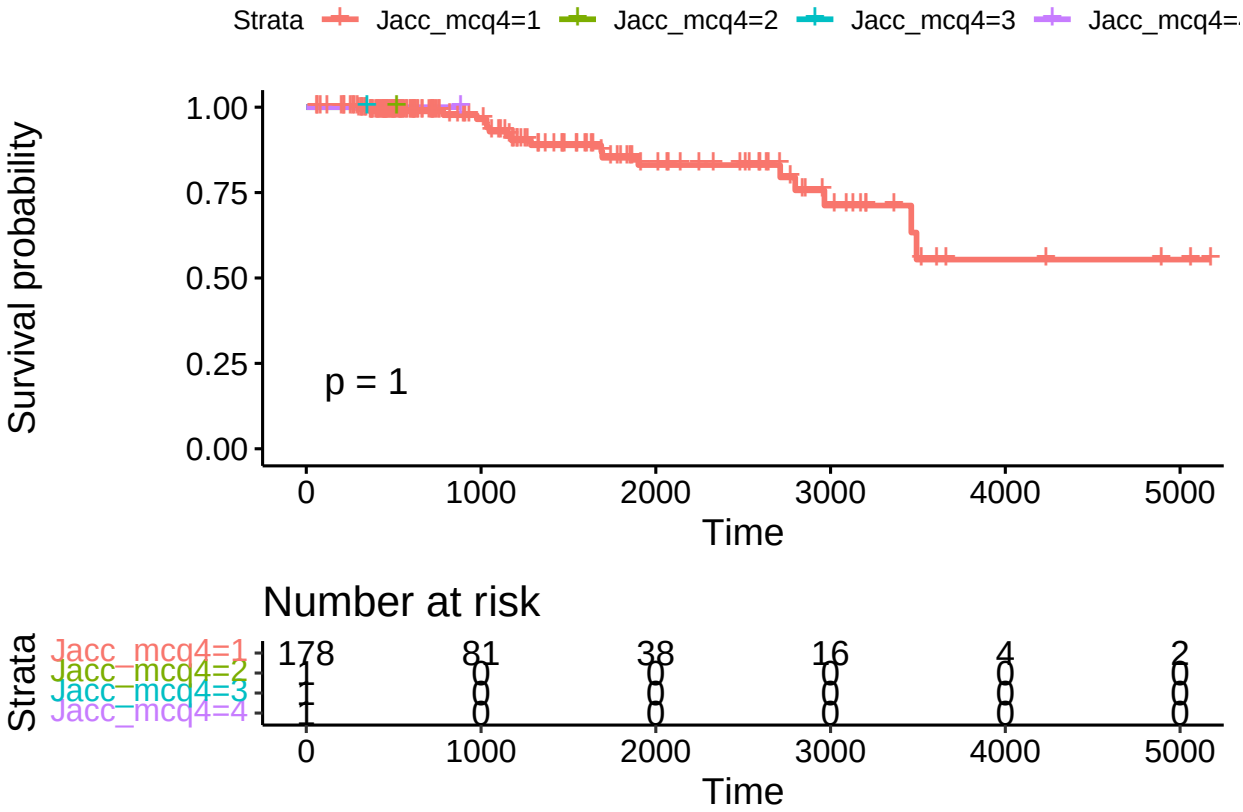
```
ggsurvplot(HC19,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS19,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

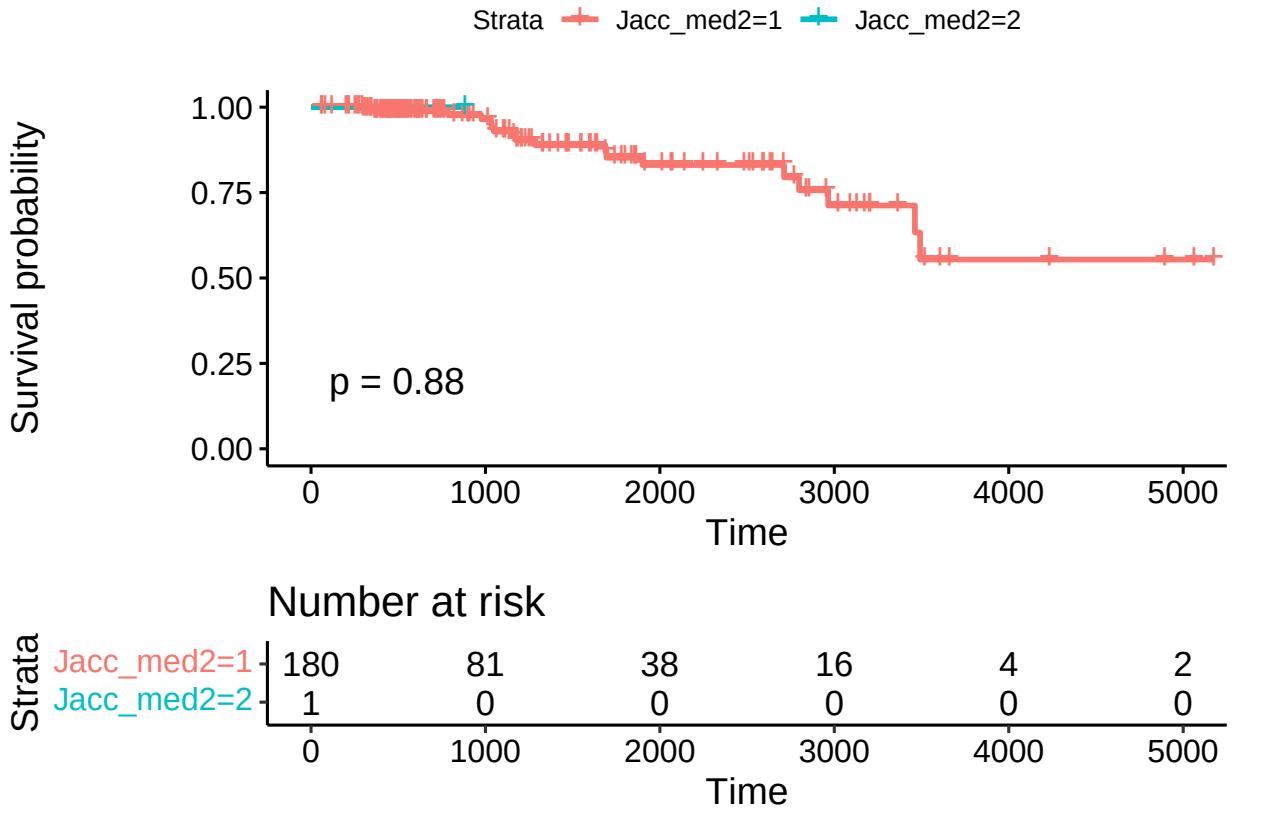



```
ggsurvplot(HW19,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

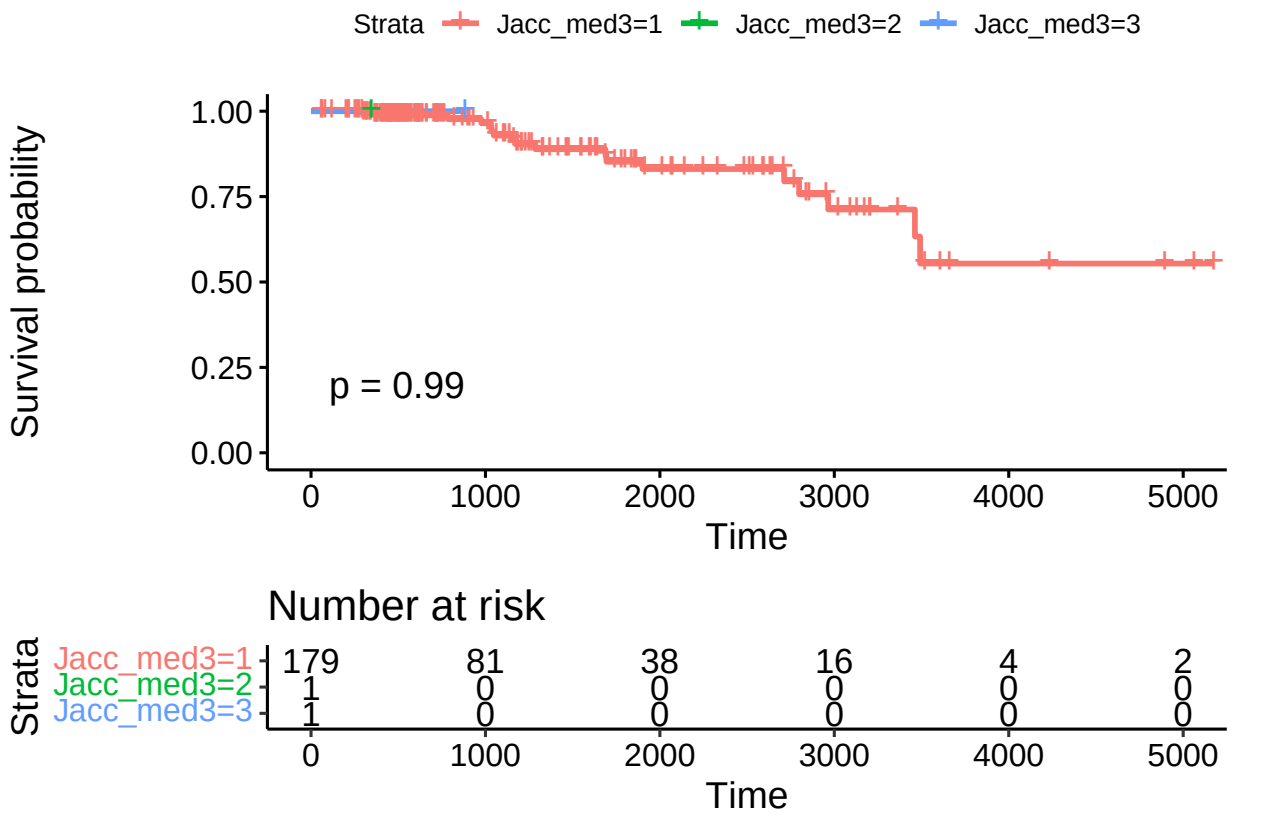


4.2.20. Jaccard distance measurement + Median Linkage

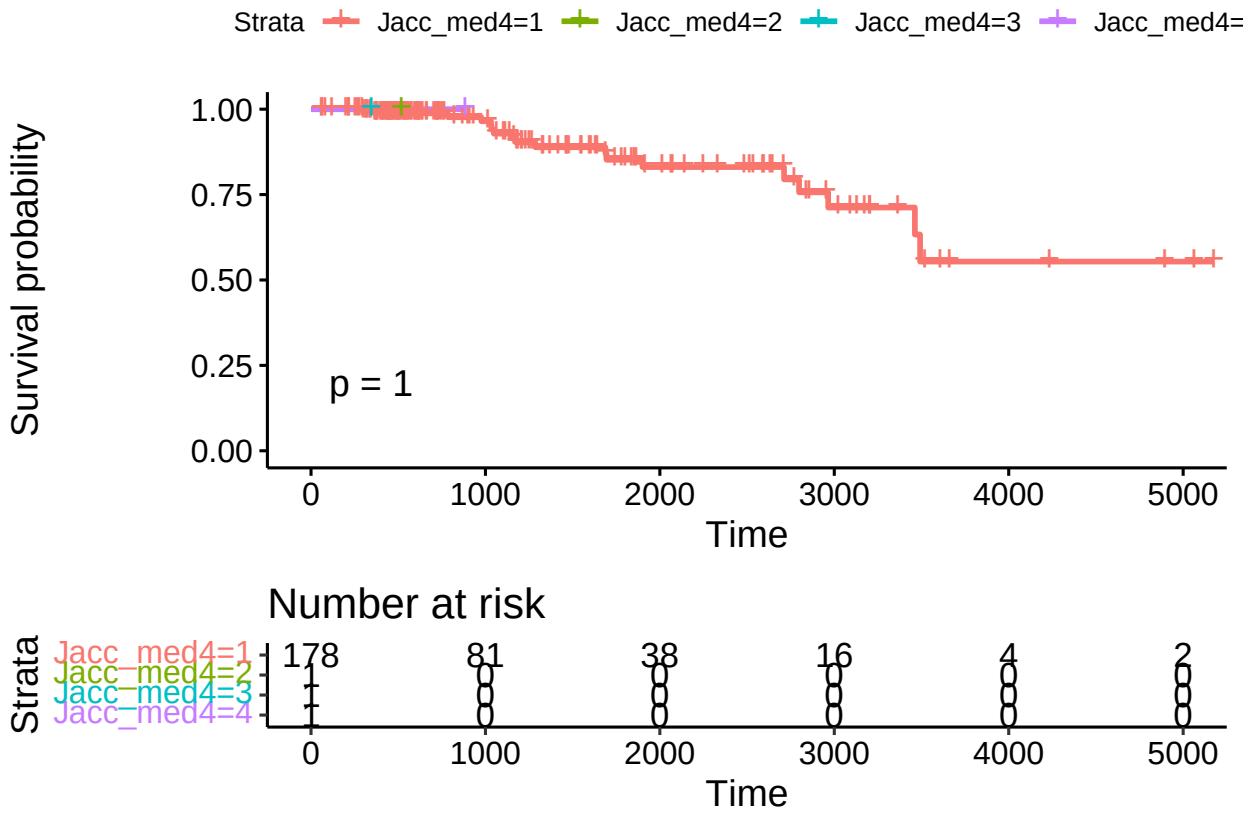
```
ggsurvplot(HC20,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS20,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

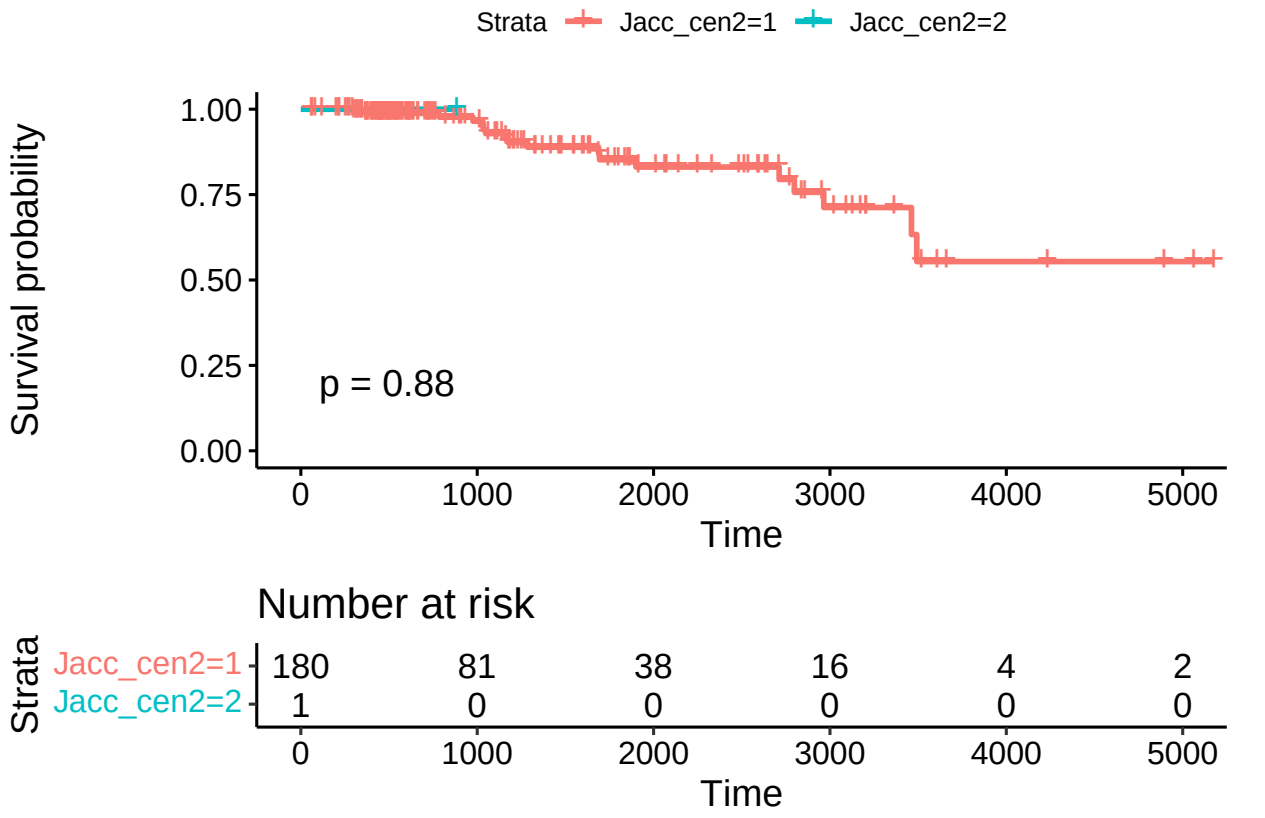



```
ggsurvplot(HW20,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

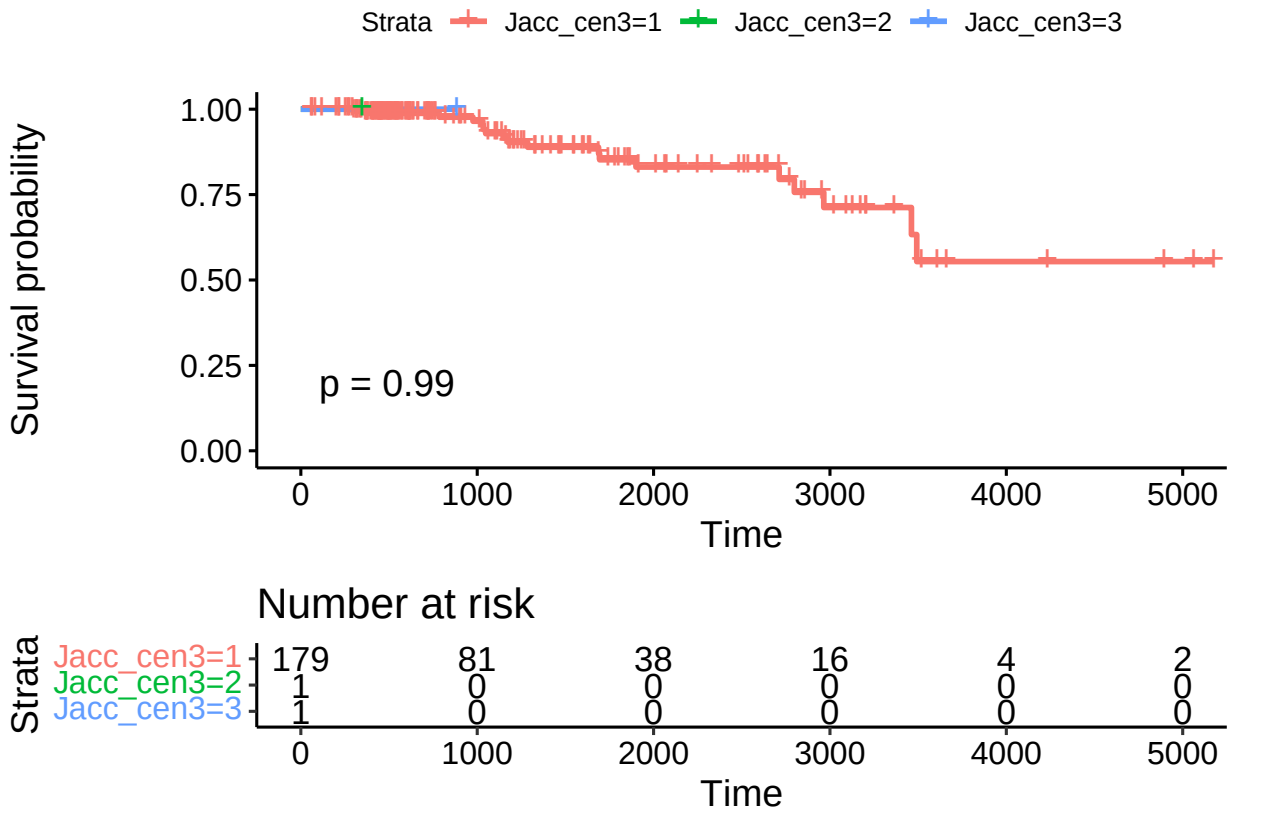


4.2.21. Jaccard distance measurement + Centroid Linkage

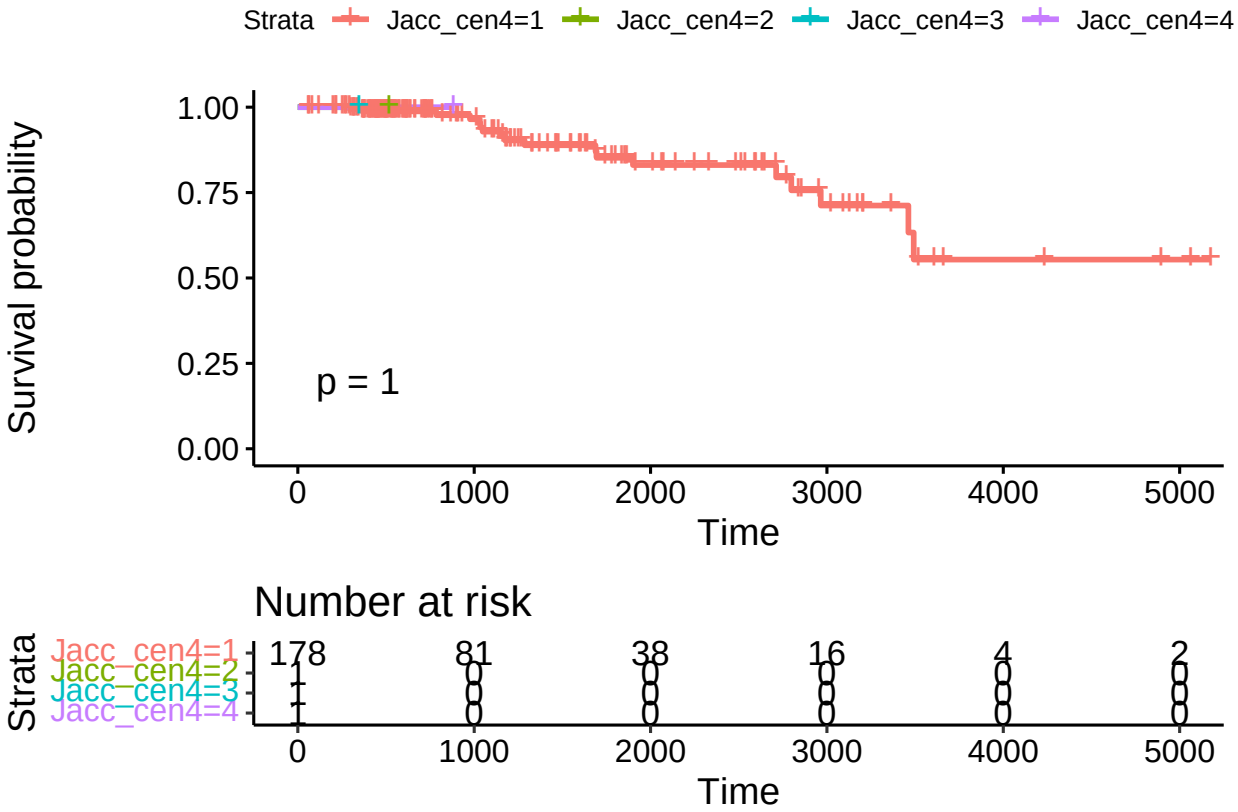
```
ggsurvplot(HC21,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS21,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

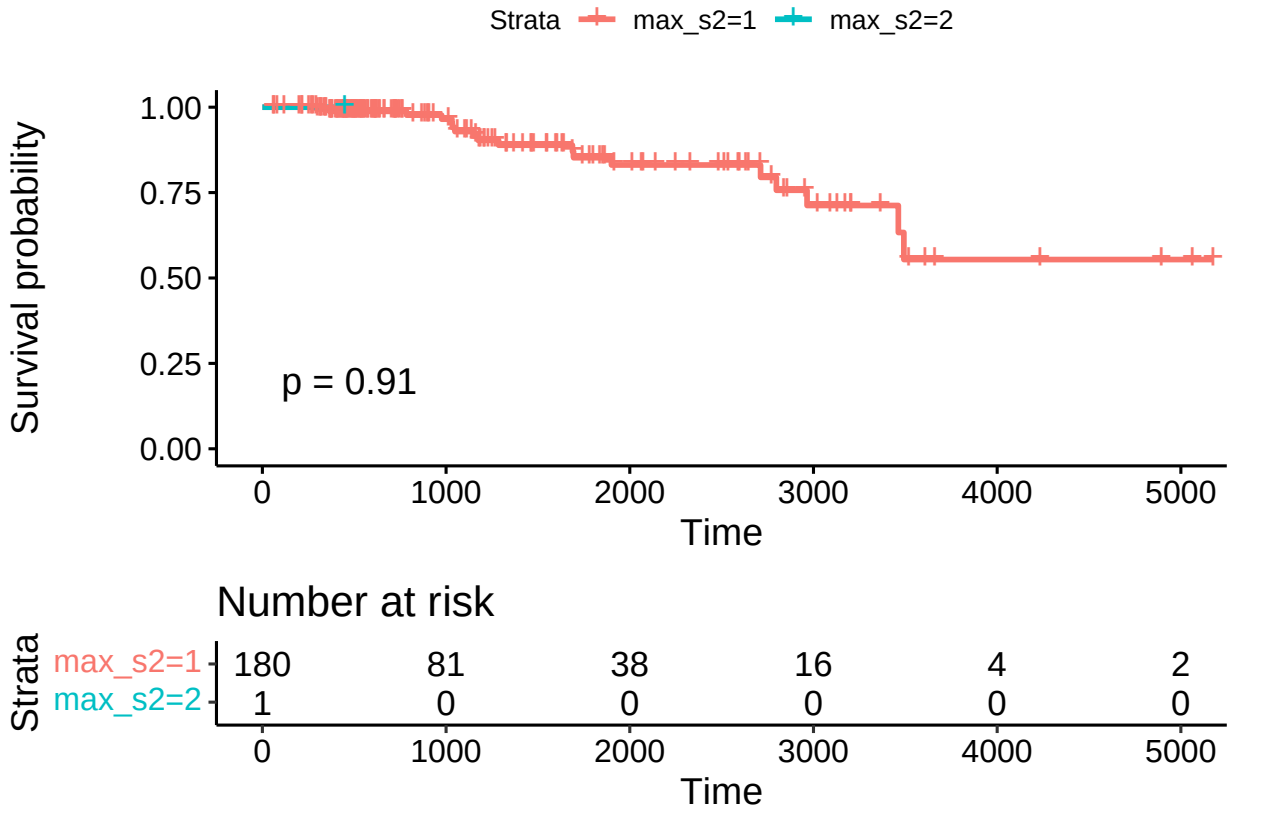



```
ggsurvplot(HW21,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

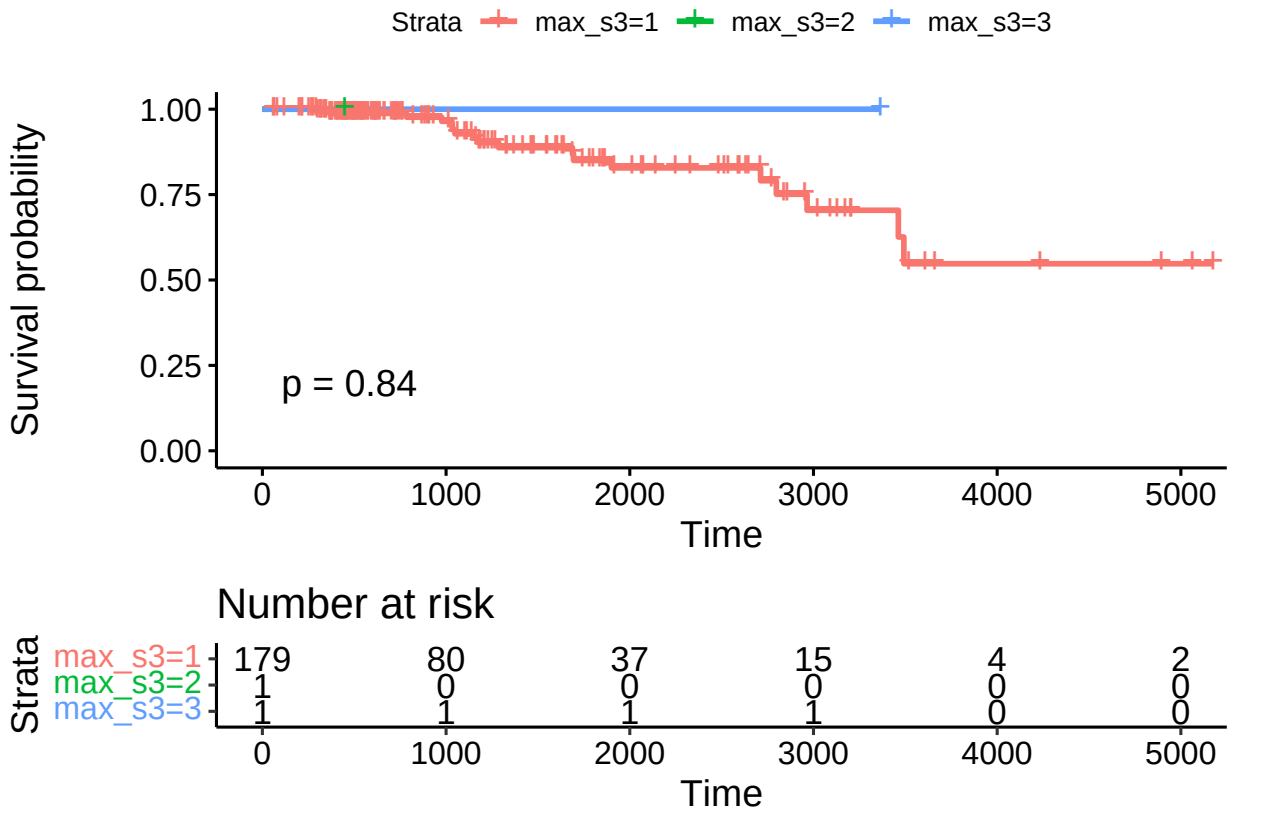


4.2.22. Maximum distance measurement + Single Linkage

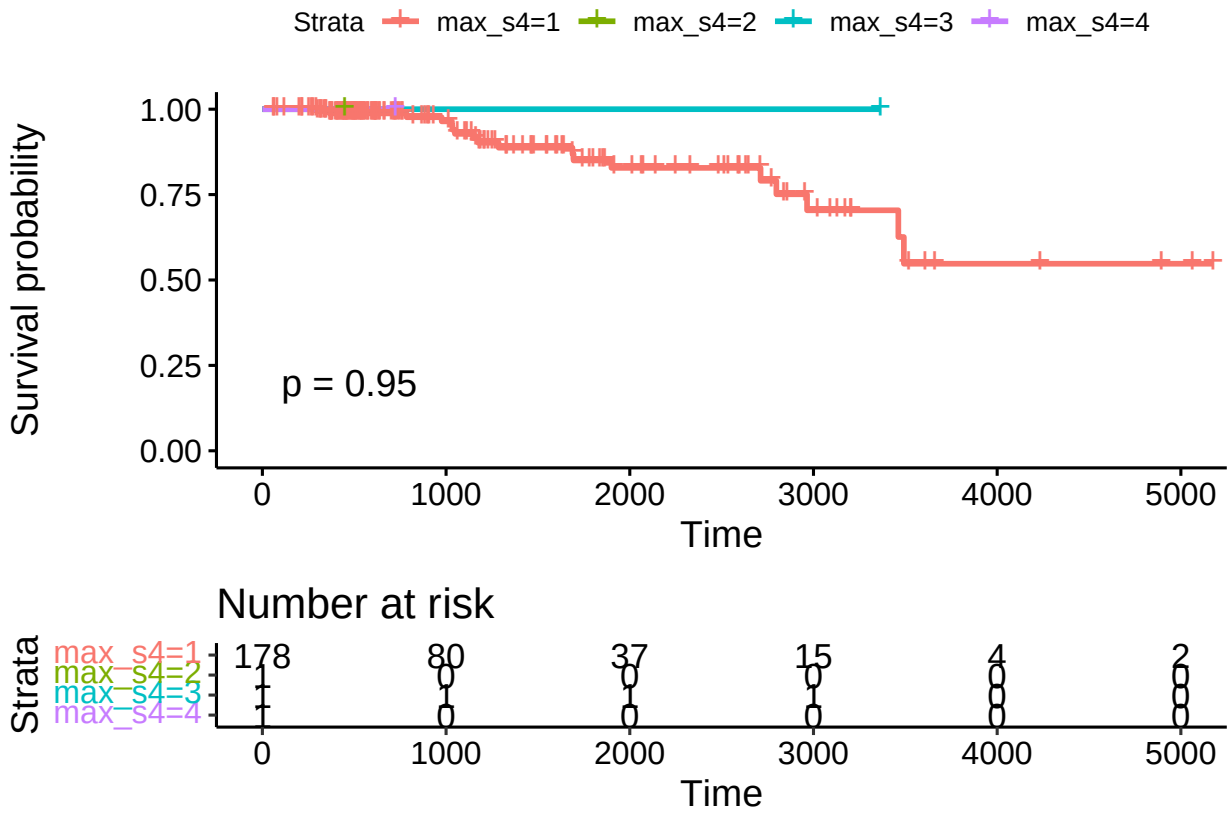
```
ggsurvplot(HC22,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS22,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

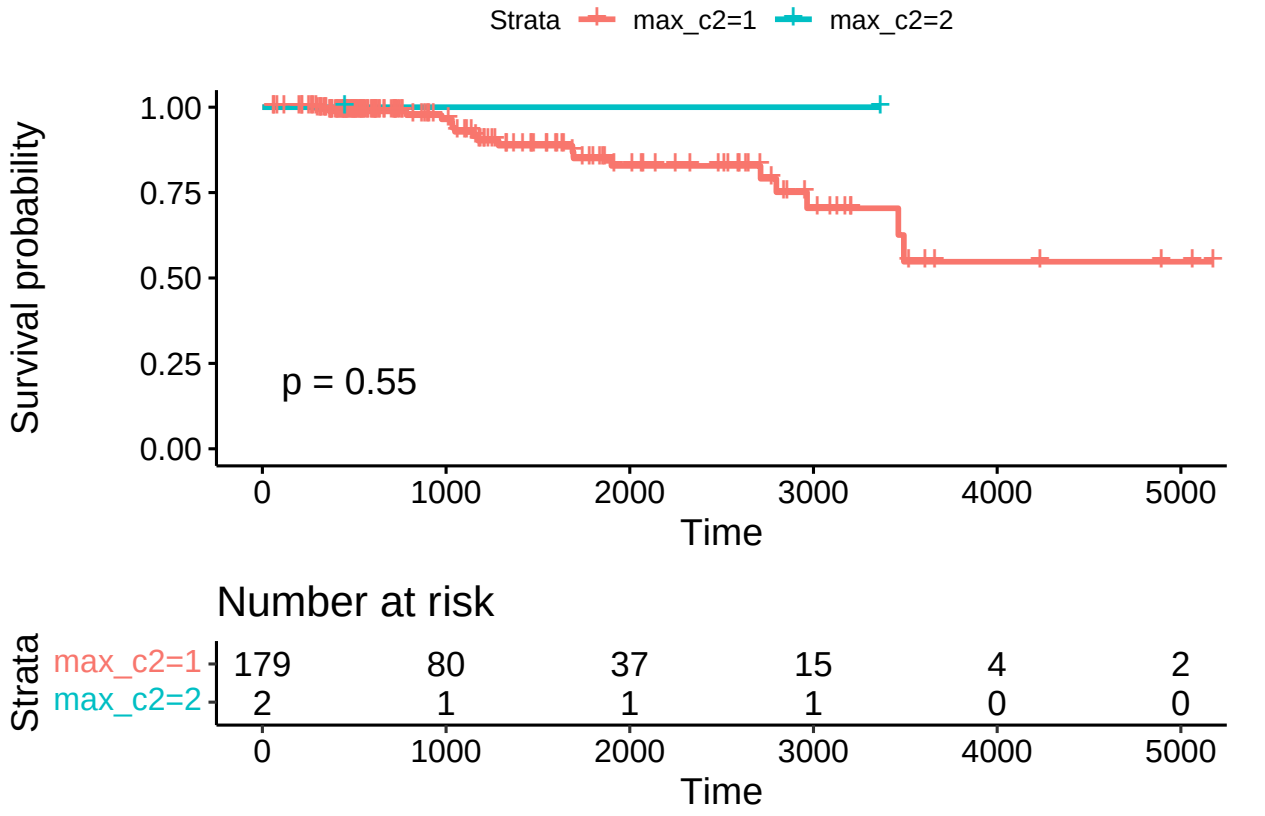



```
ggsurvplot(HW22,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

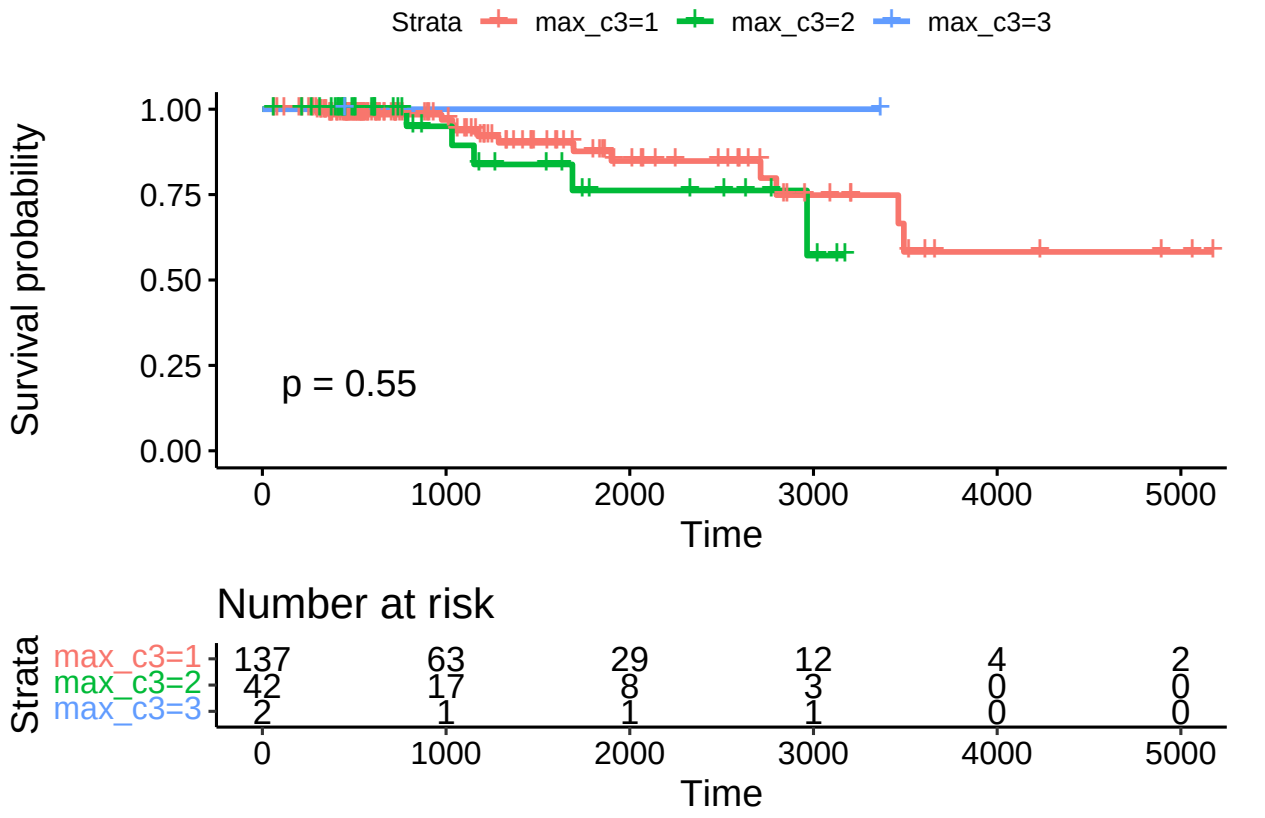


4.2.23. Maximum distance measurement + Average Linkage

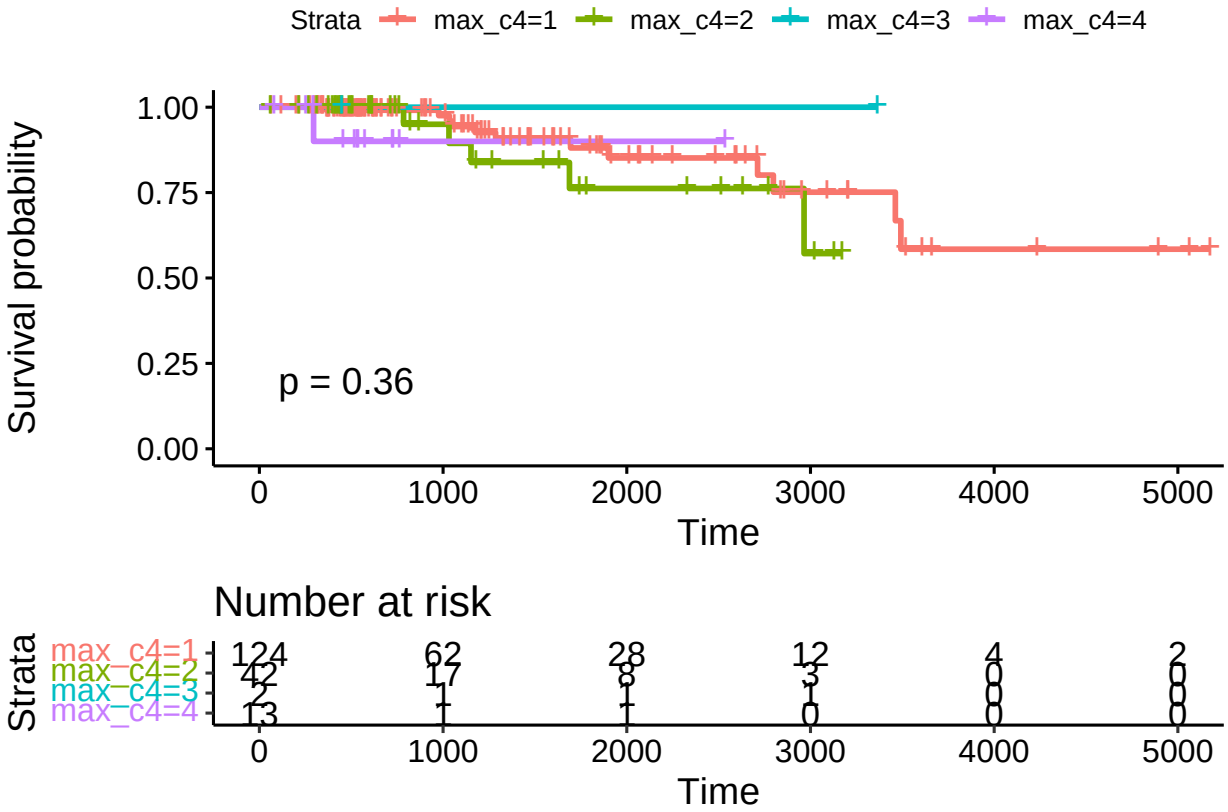
```
ggsurvplot(HC23,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS23,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

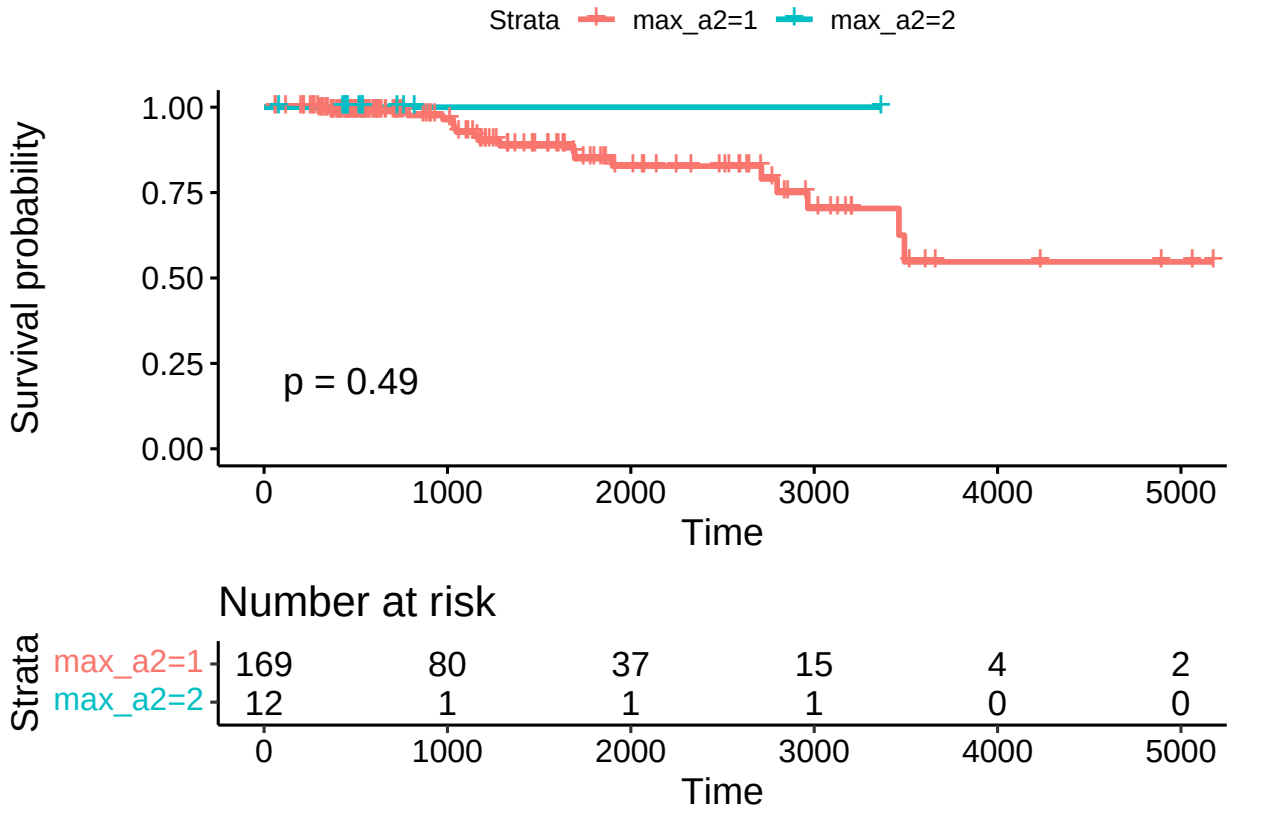



```
ggsurvplot(HW23,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

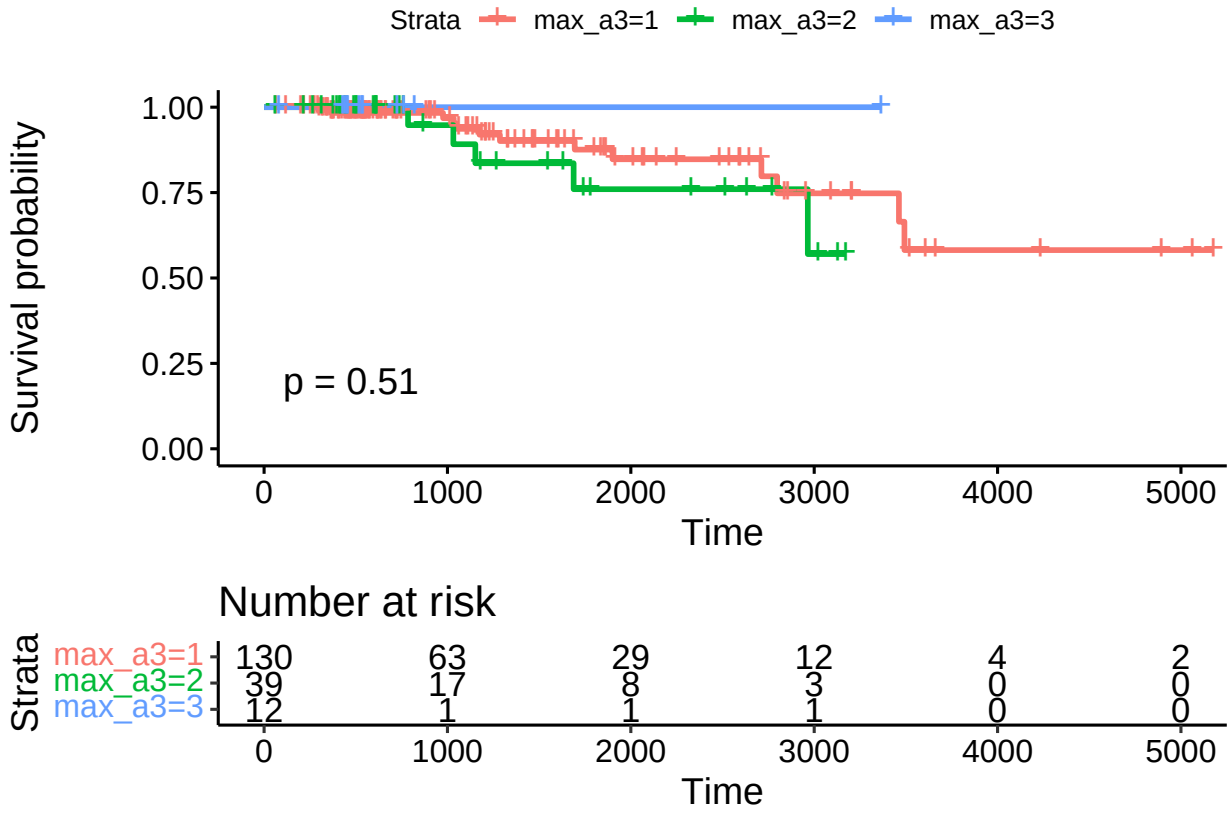


4.2.24. Maximum distance measurement + Complete Linkage

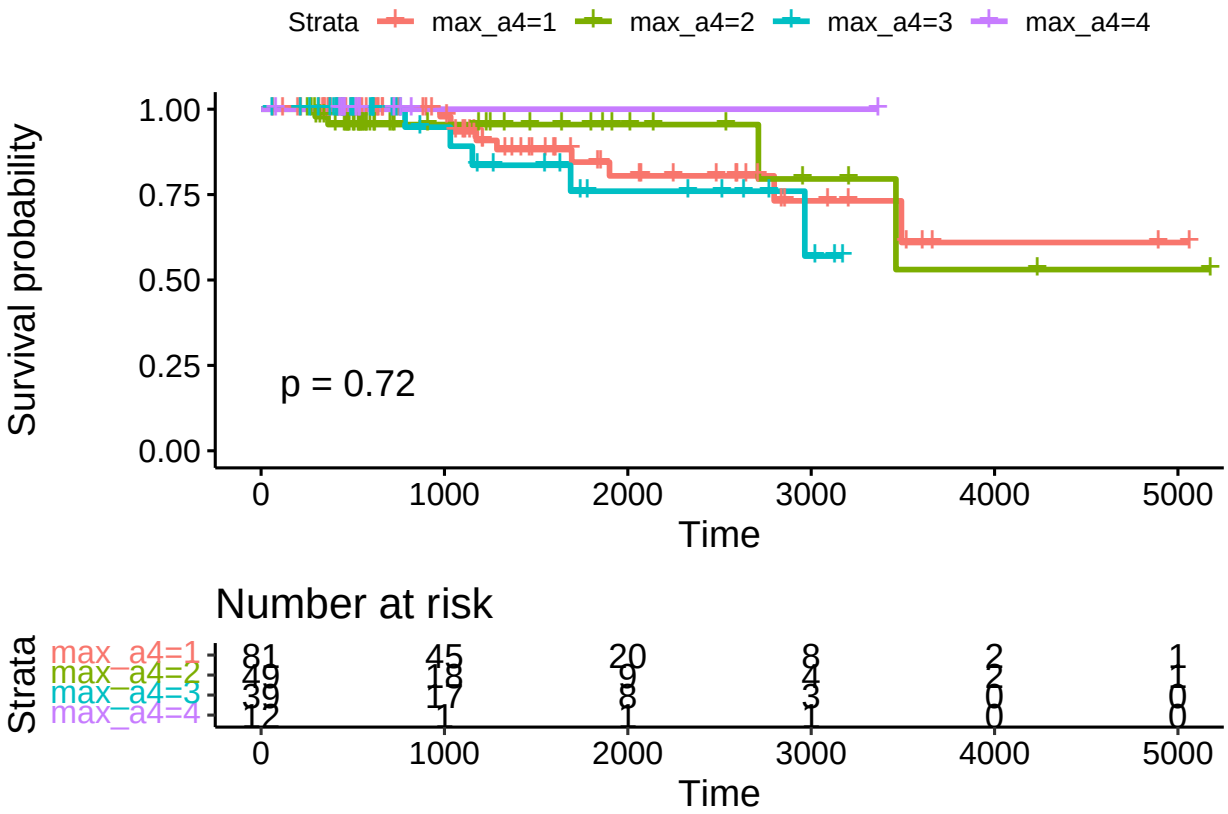
```
ggsurvplot(HC24,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS24,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

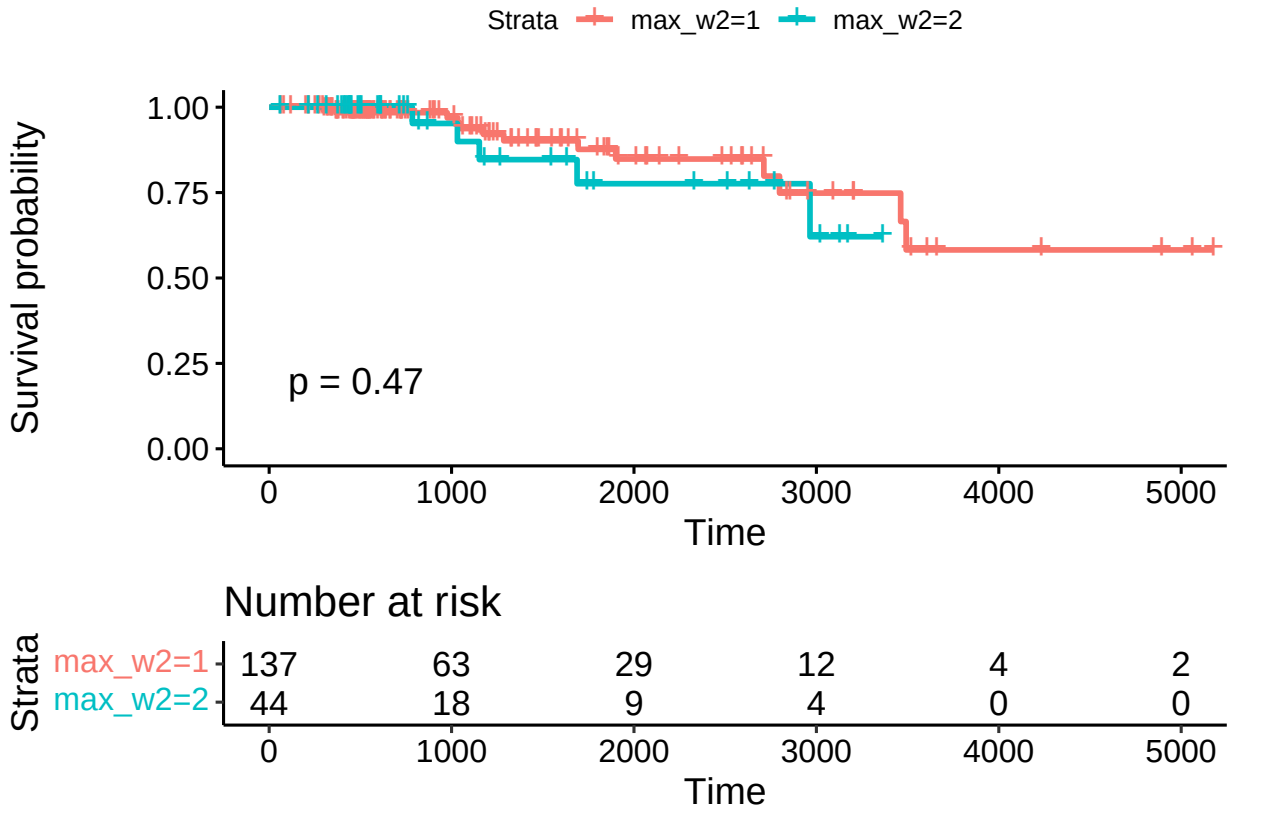



```
ggsurvplot(HW24,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

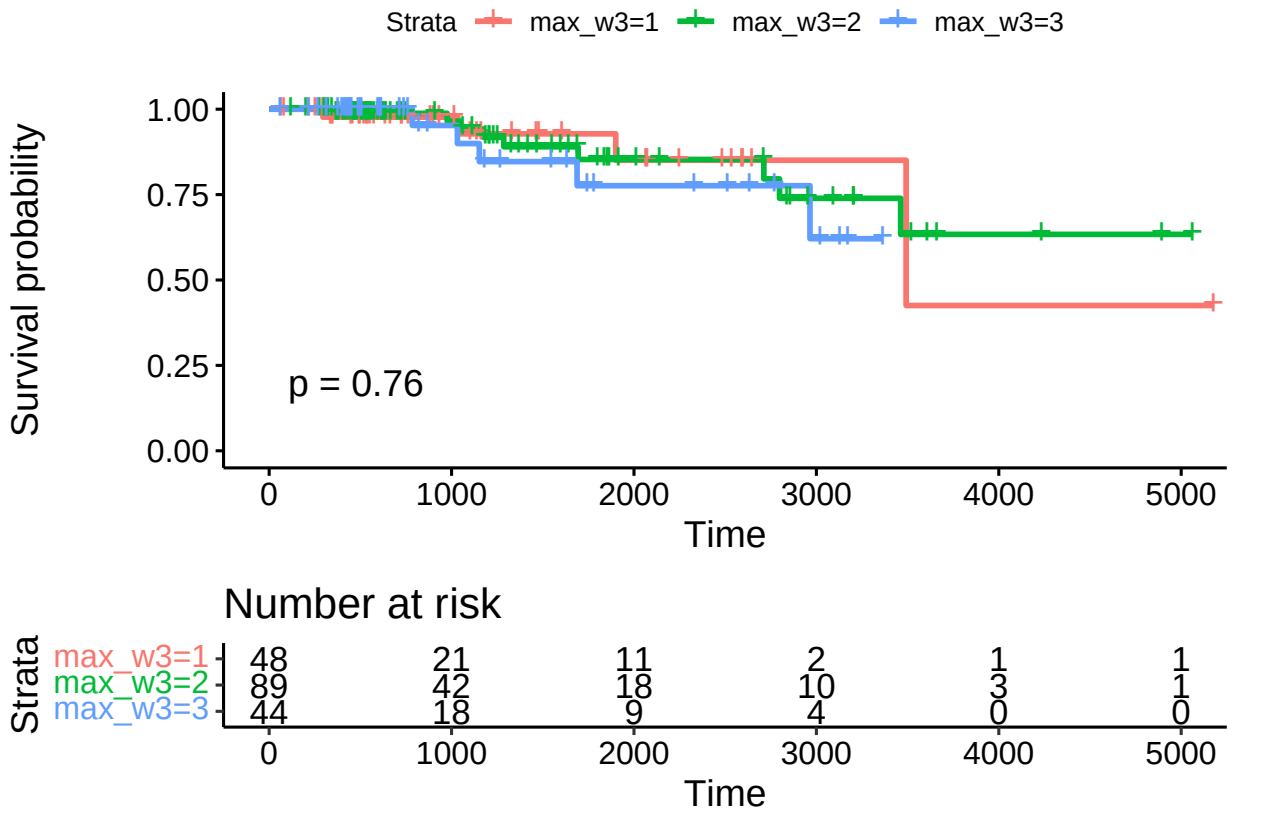


4.2.25. Maximum distance measurement + Ward Linkage

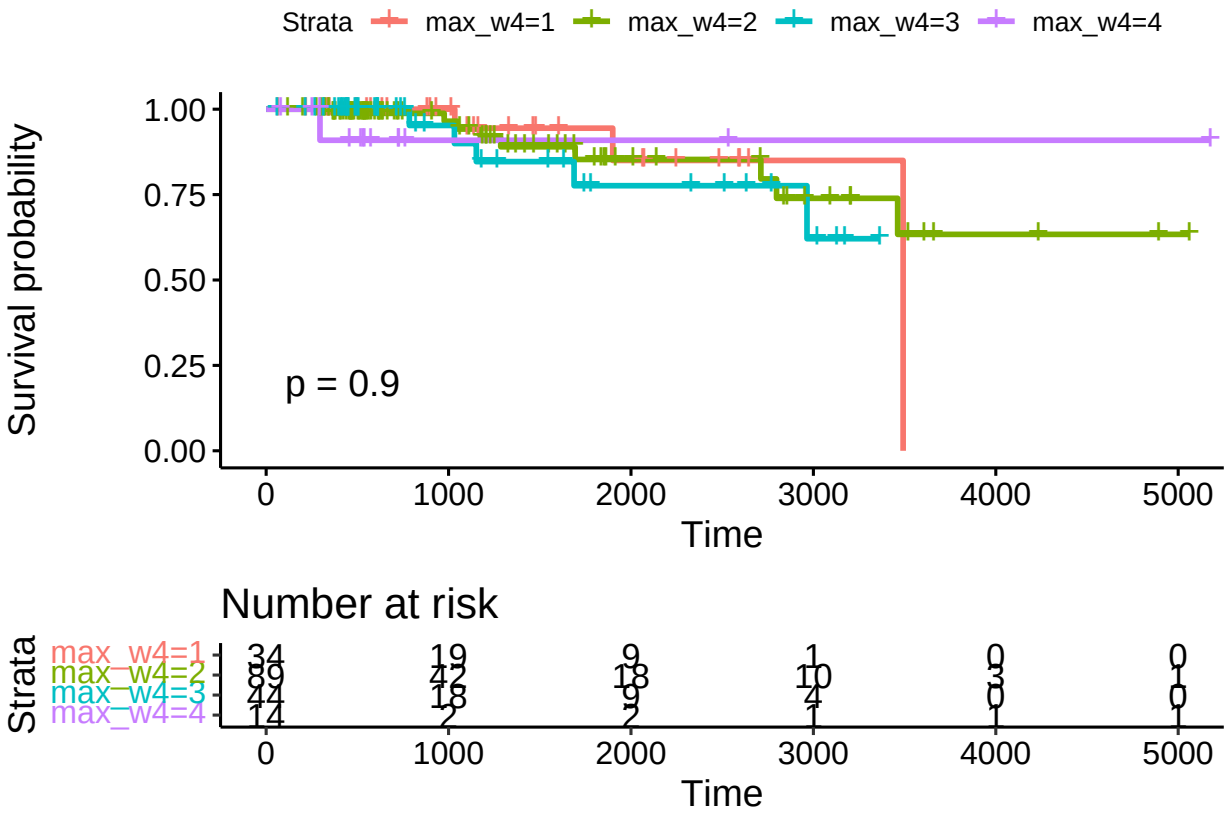
```
ggsurvplot(HC25,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS25,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

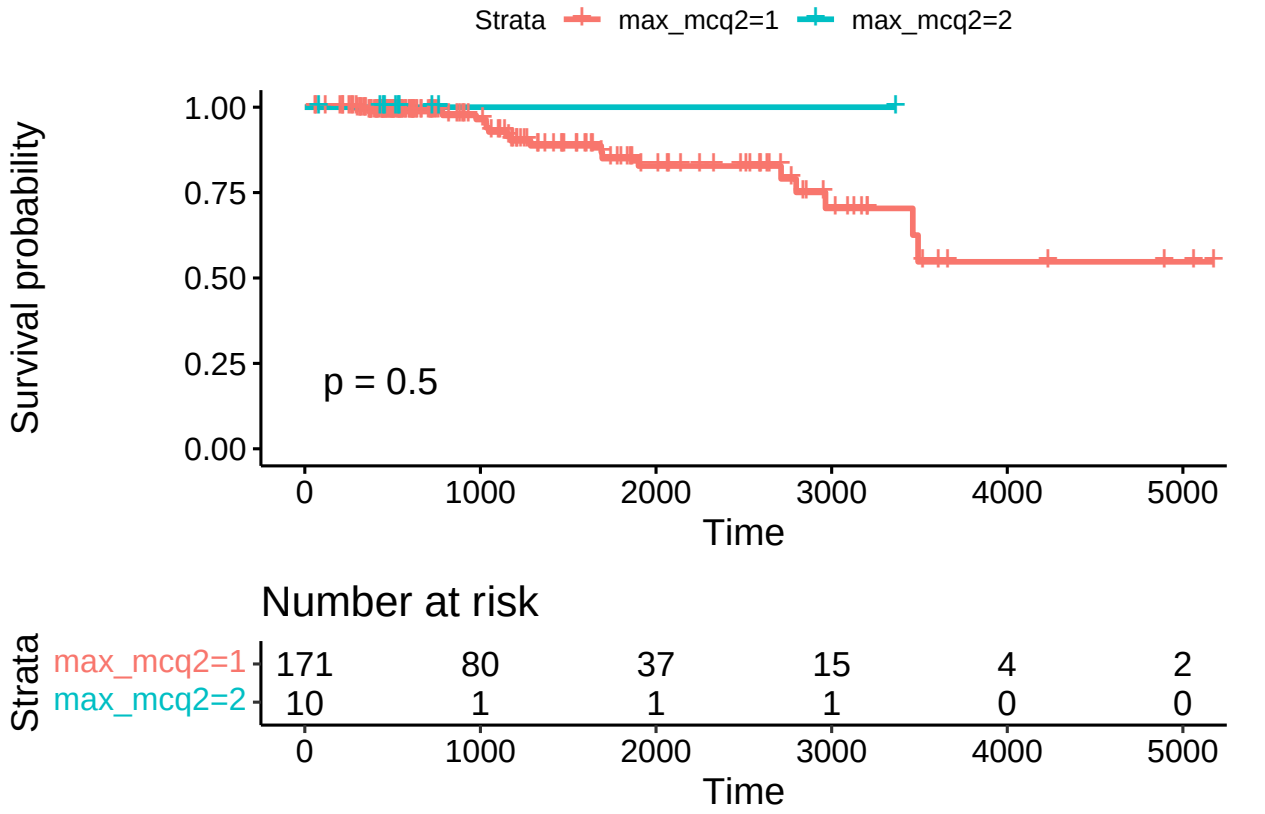



```
ggsurvplot(HW25,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

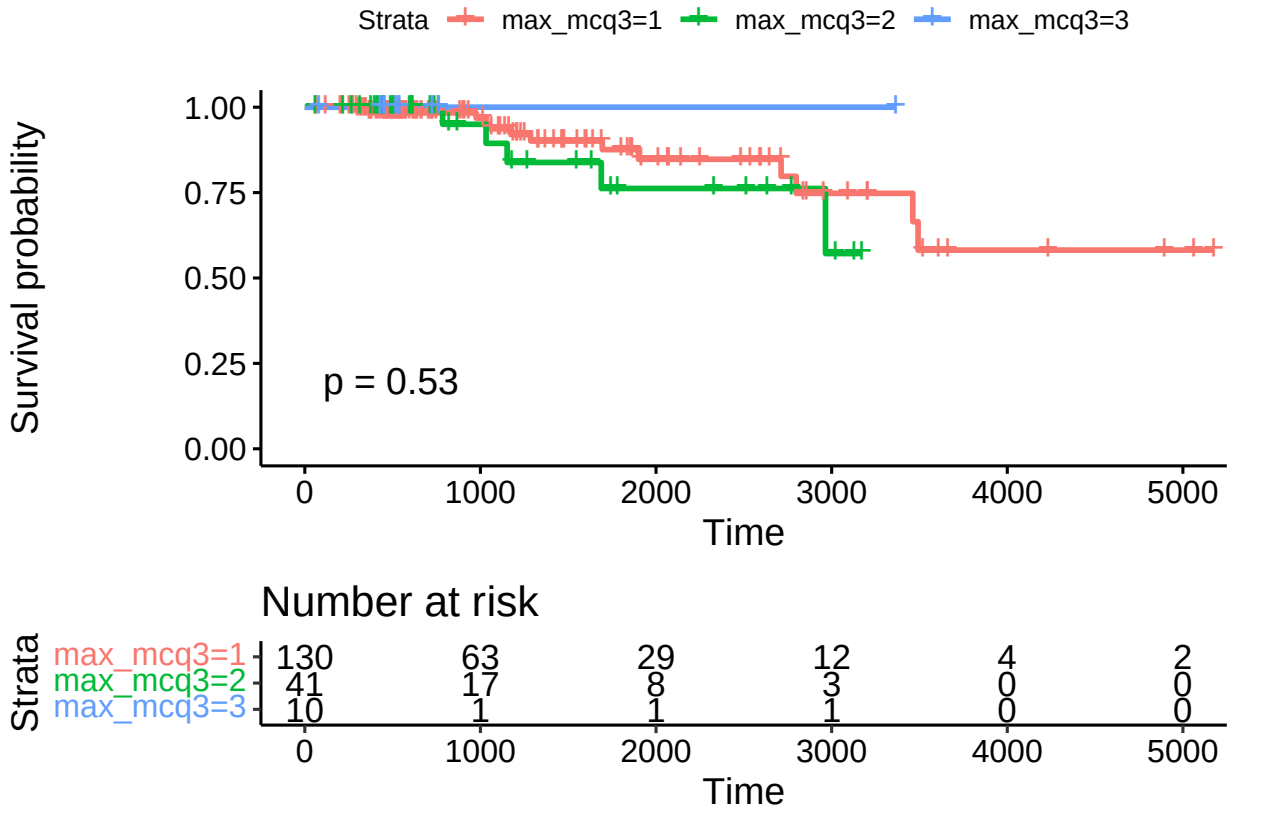


4.2.26. Maximum distance measurement + Mcquitty Linkage

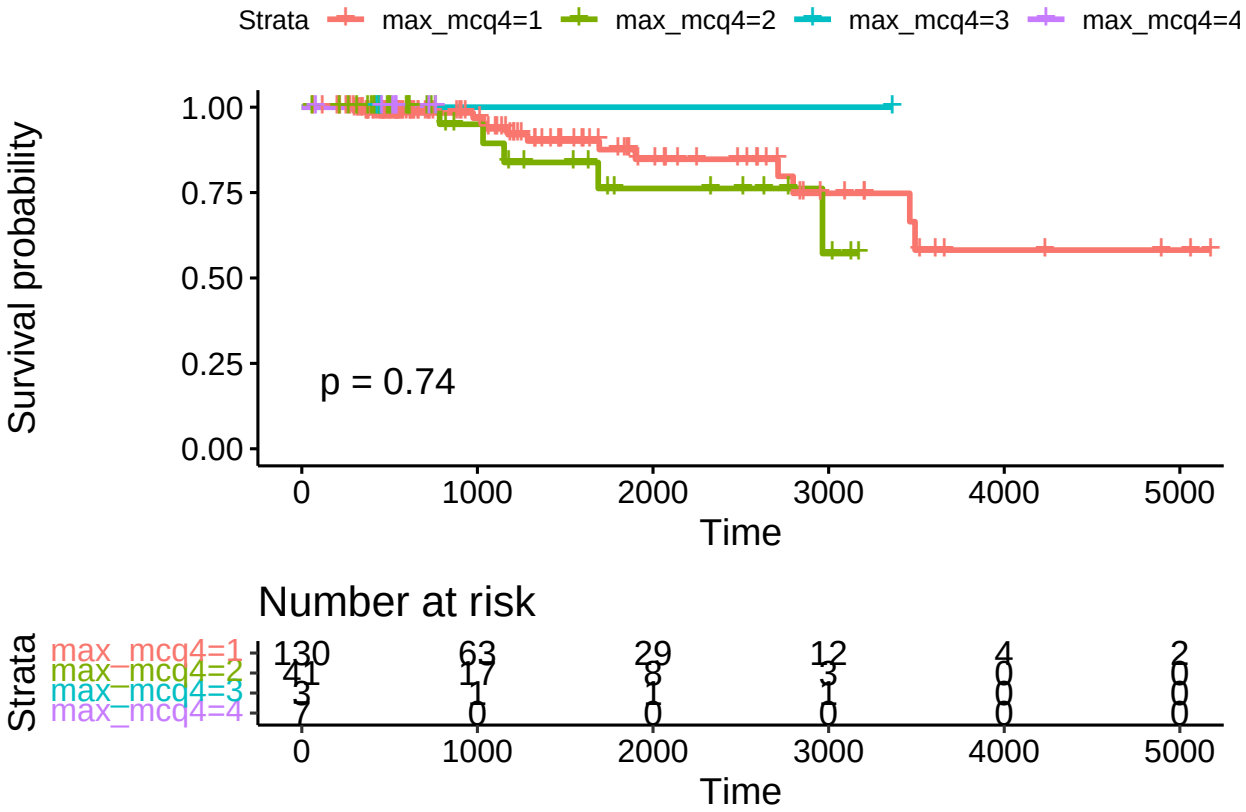
```
ggsurvplot(HC26,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS26,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

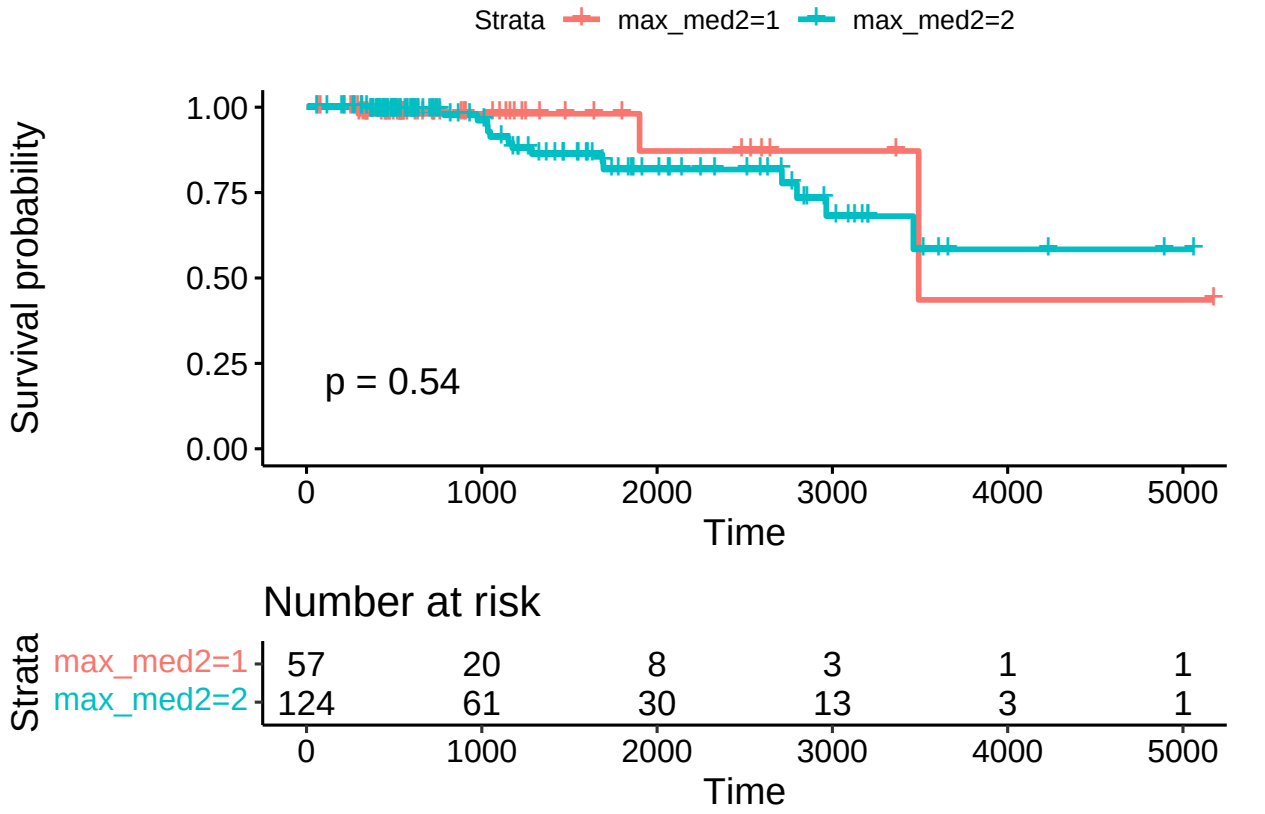



```
ggsurvplot(HW26,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

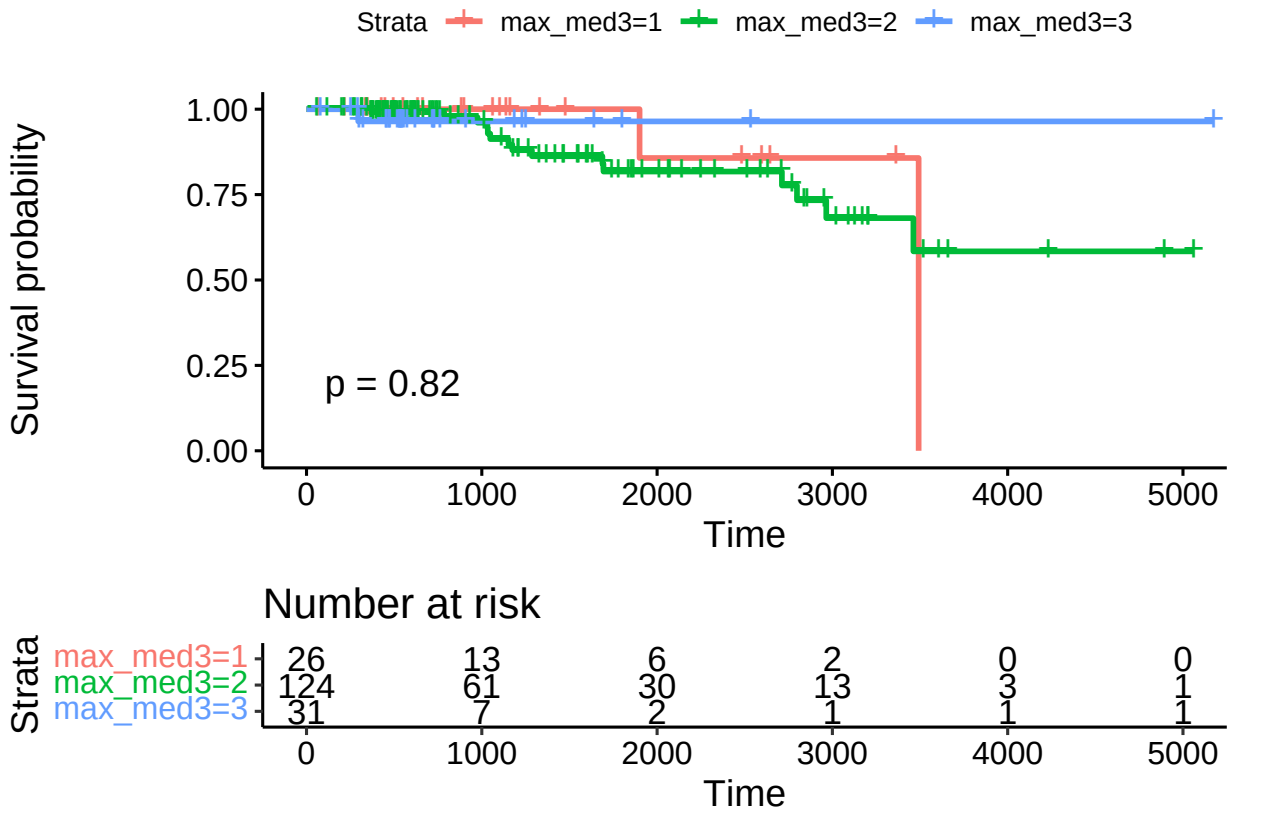


4.2.27. Maximum distance measurement + Median Linkage

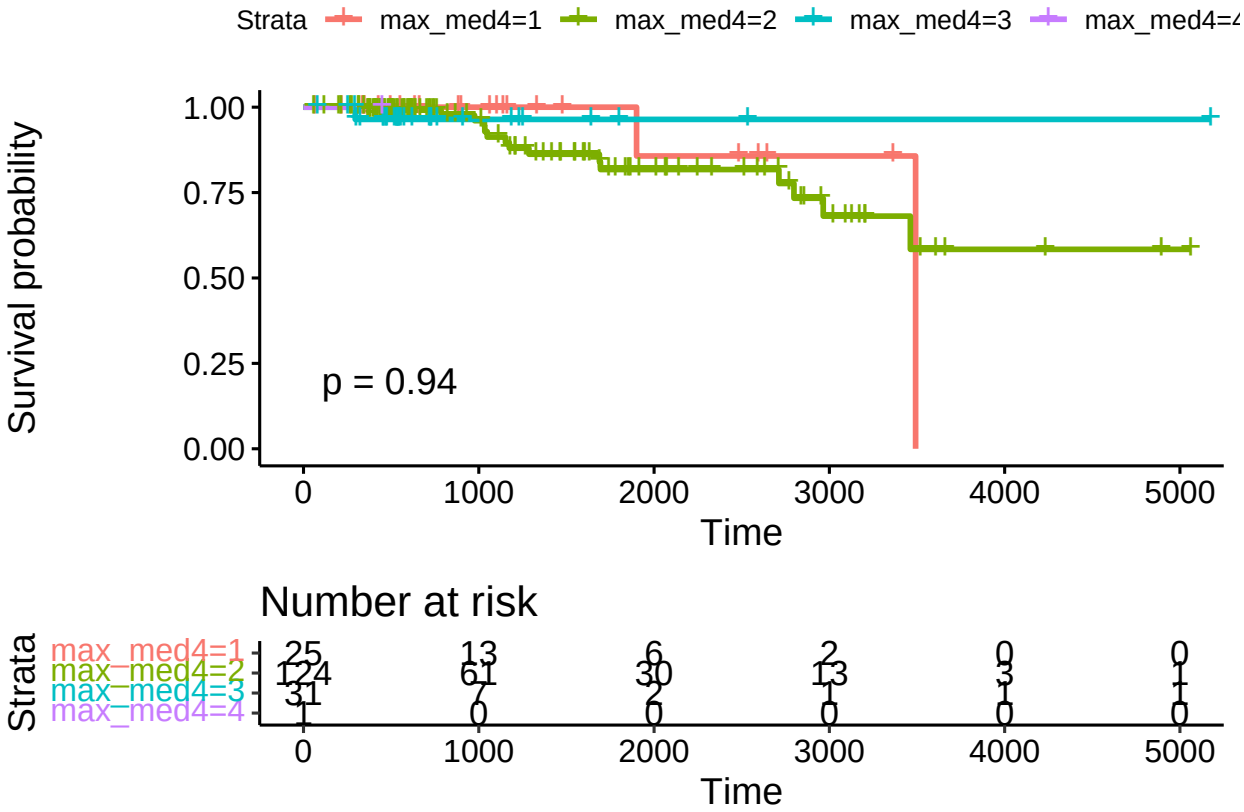
```
ggsurvplot(HC27,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS27,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

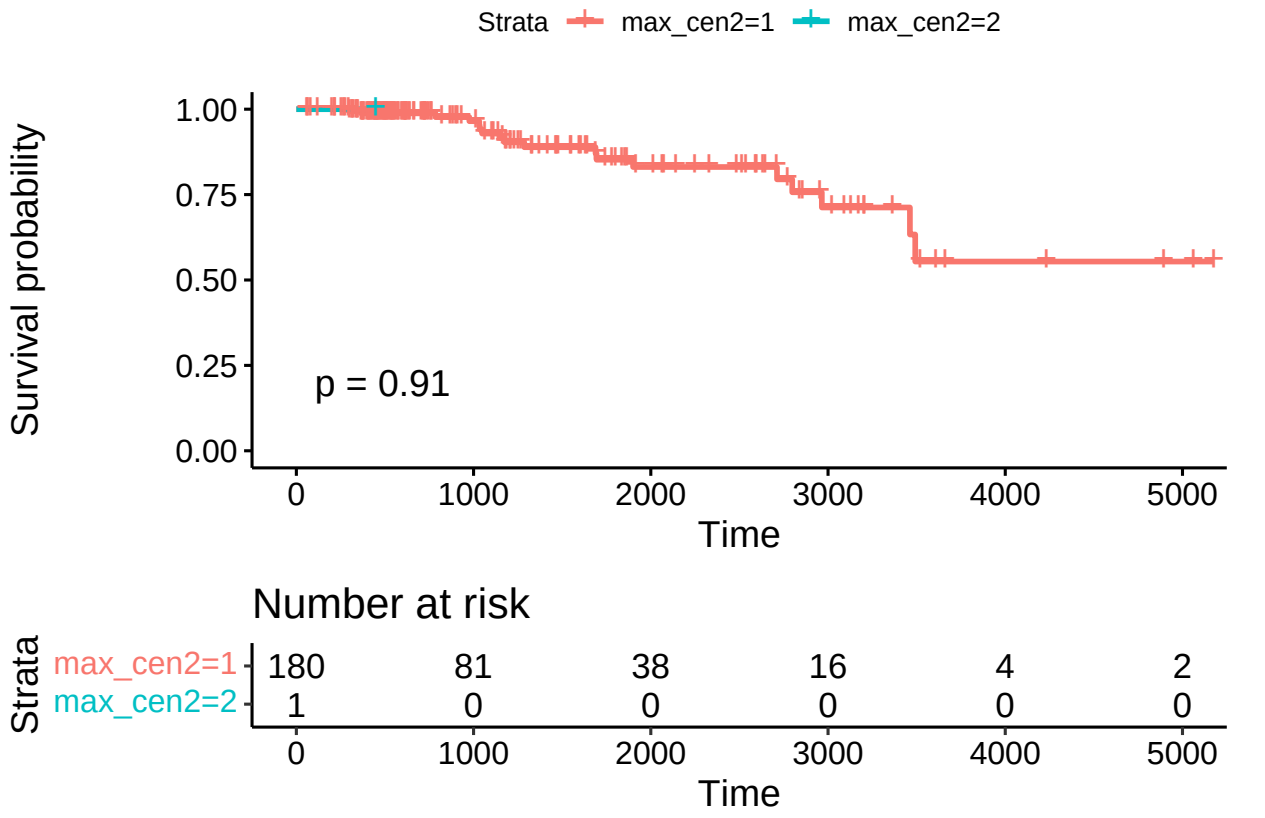



```
ggsurvplot(HW27,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

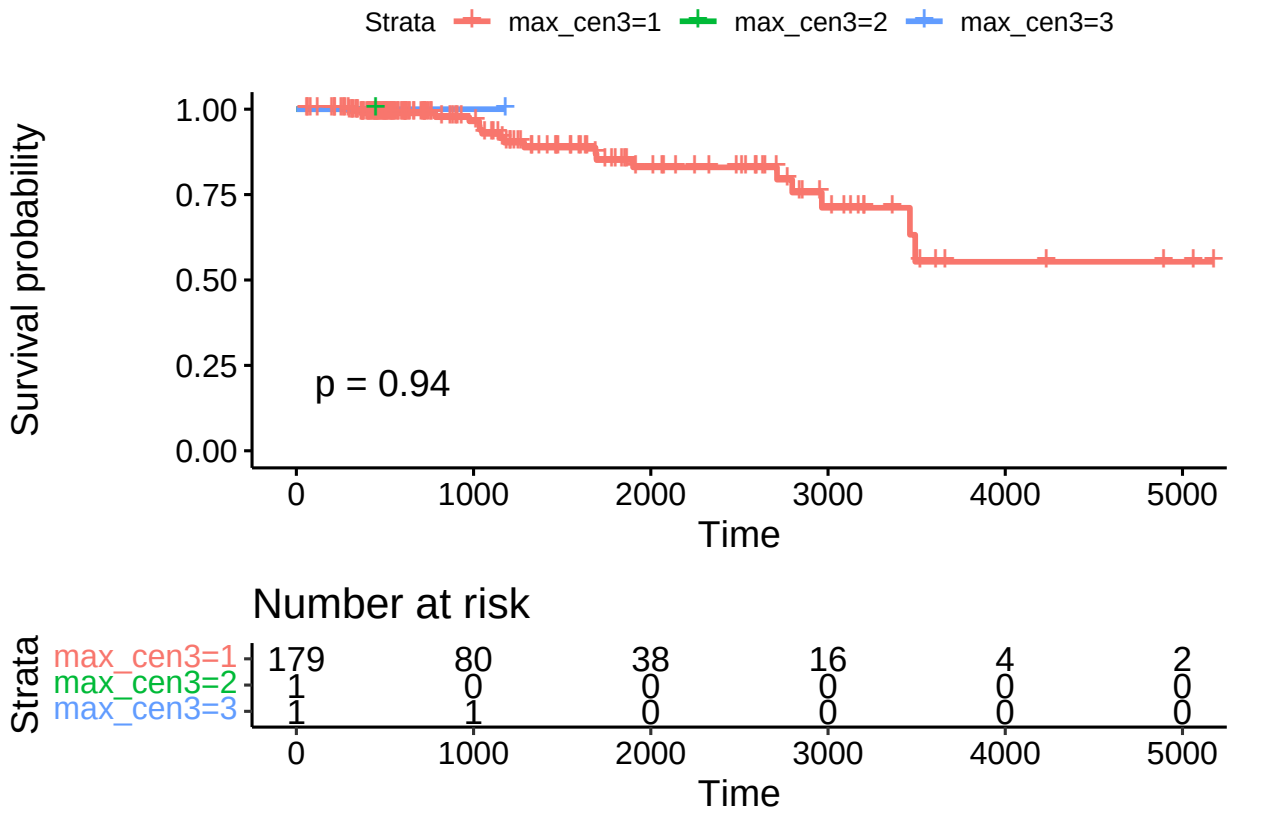


4.2.28. Maximum distance measurement + Centroid Linkage

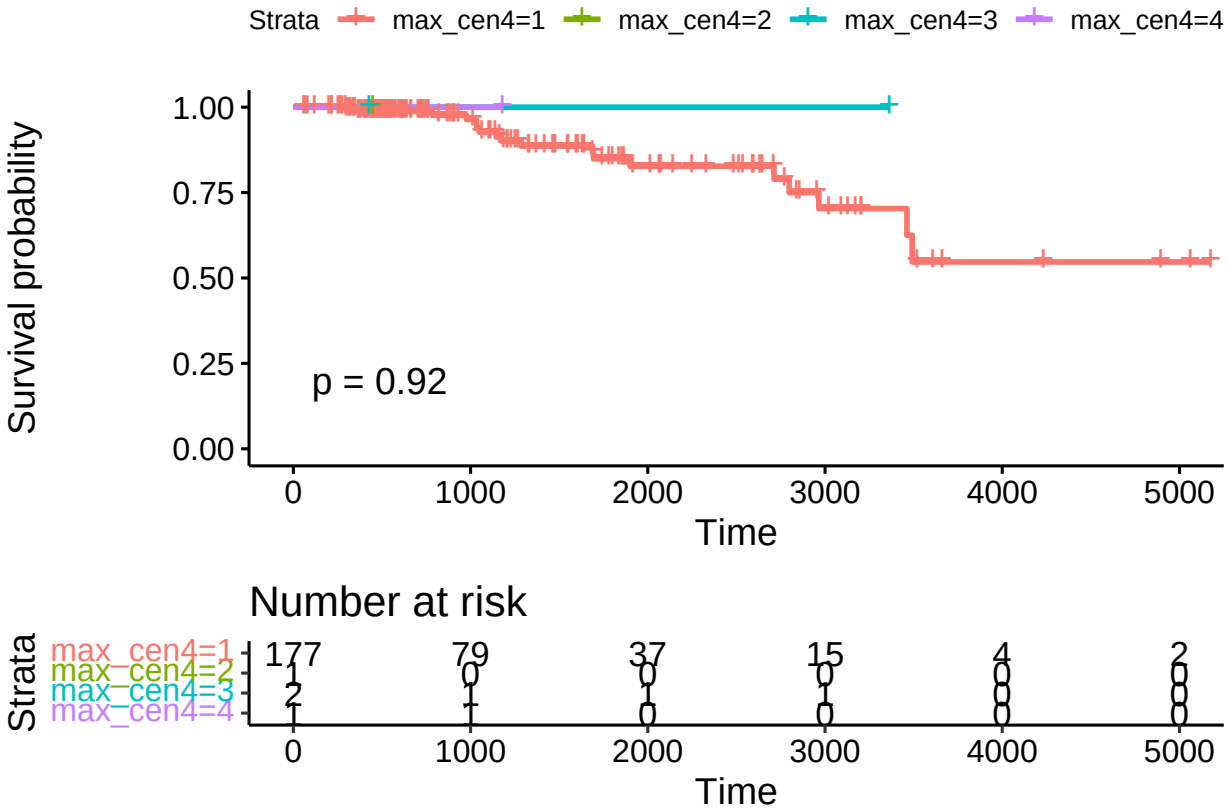
```
ggsurvplot(HC28,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS28,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

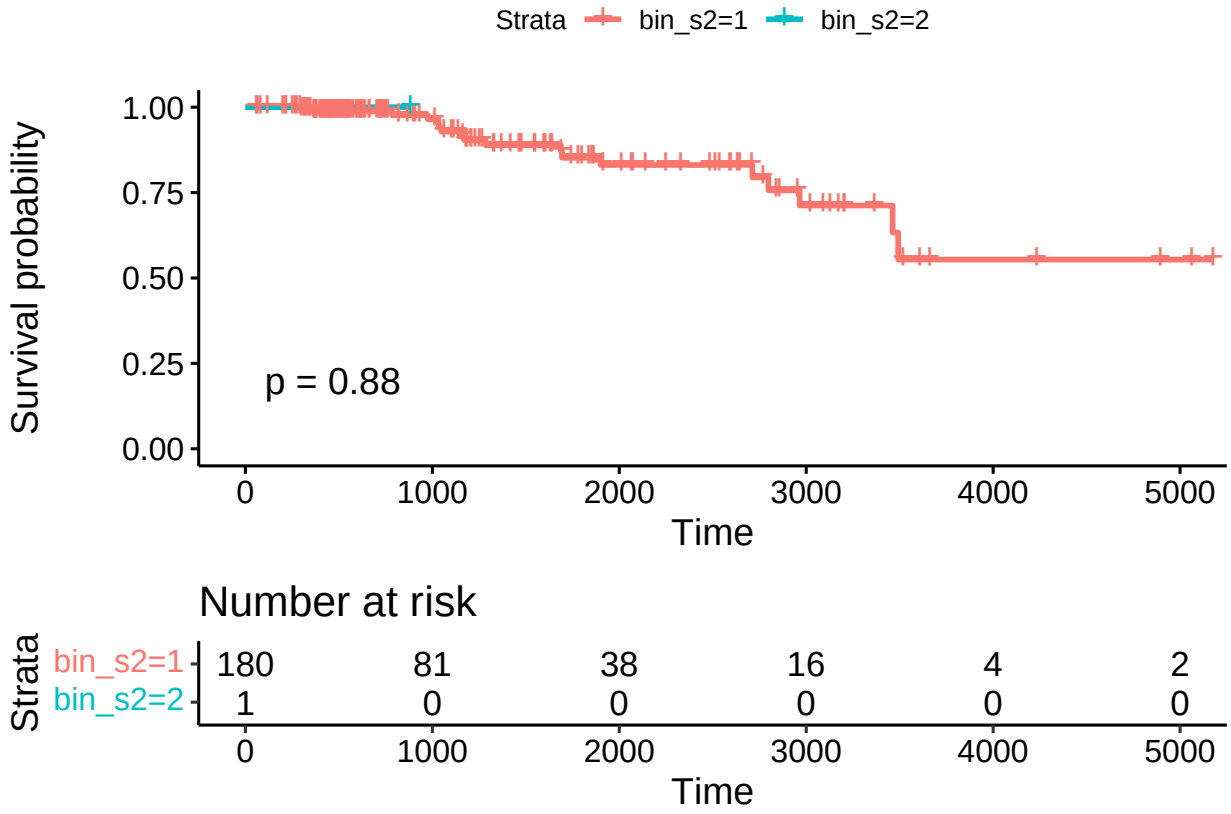



```
ggsurvplot(HW28,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

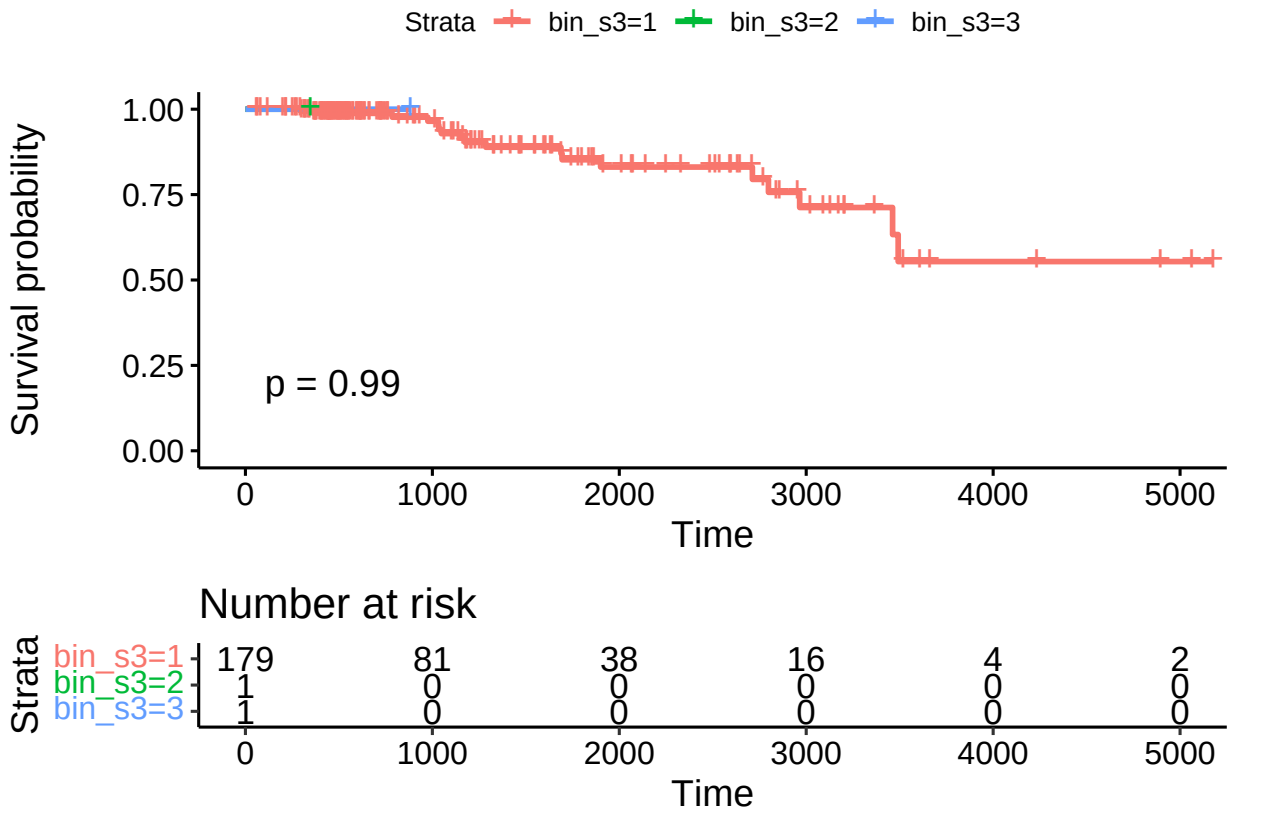


4.2.29. Binary distance measurement + Single Linkage

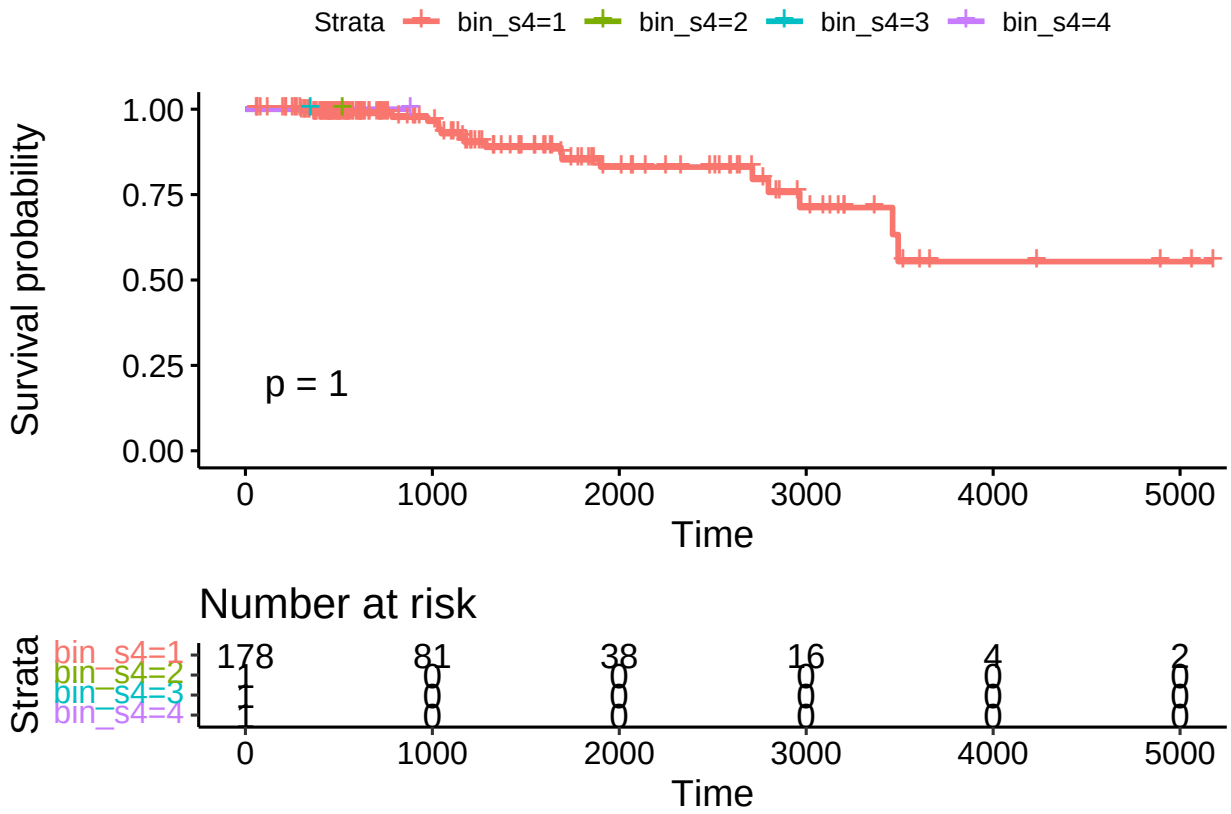
```
ggsurvplot(HC29,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS29,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

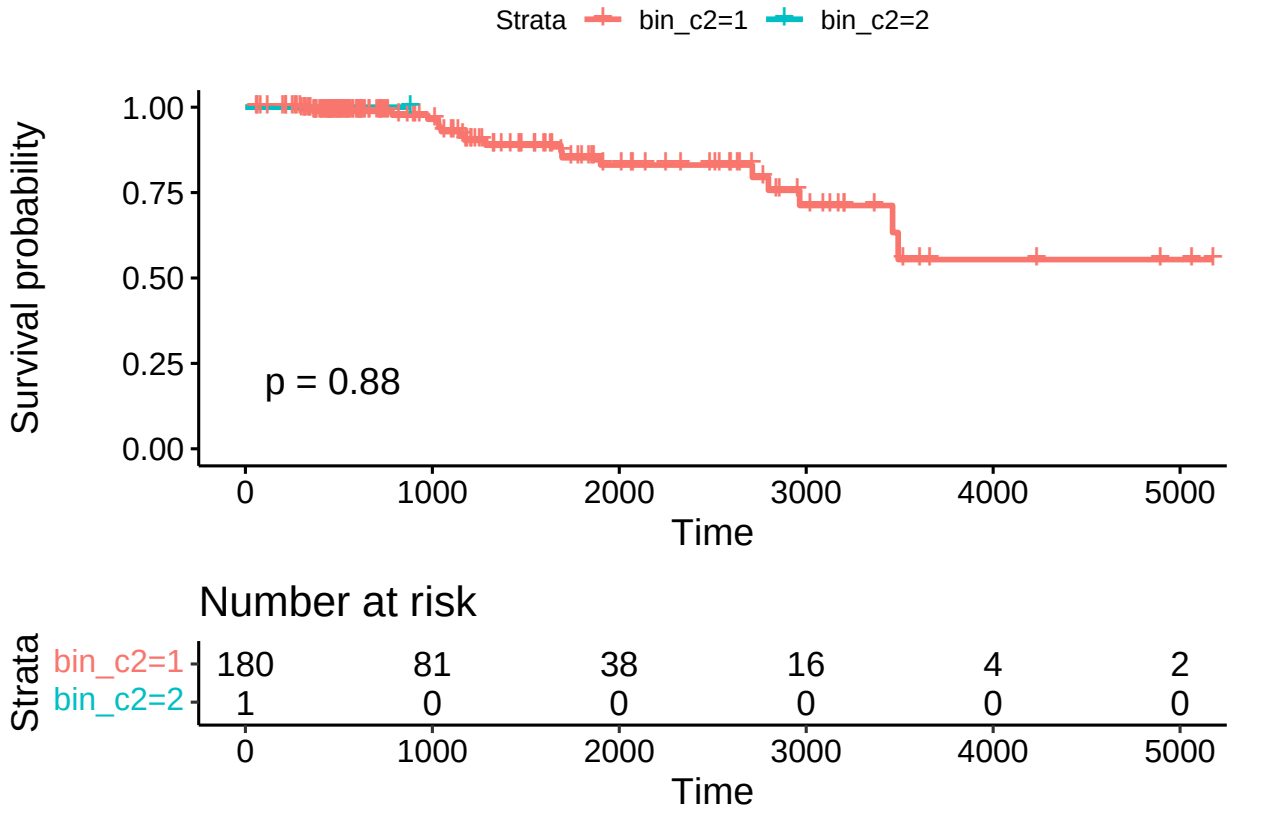



```
ggsurvplot(HW29,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

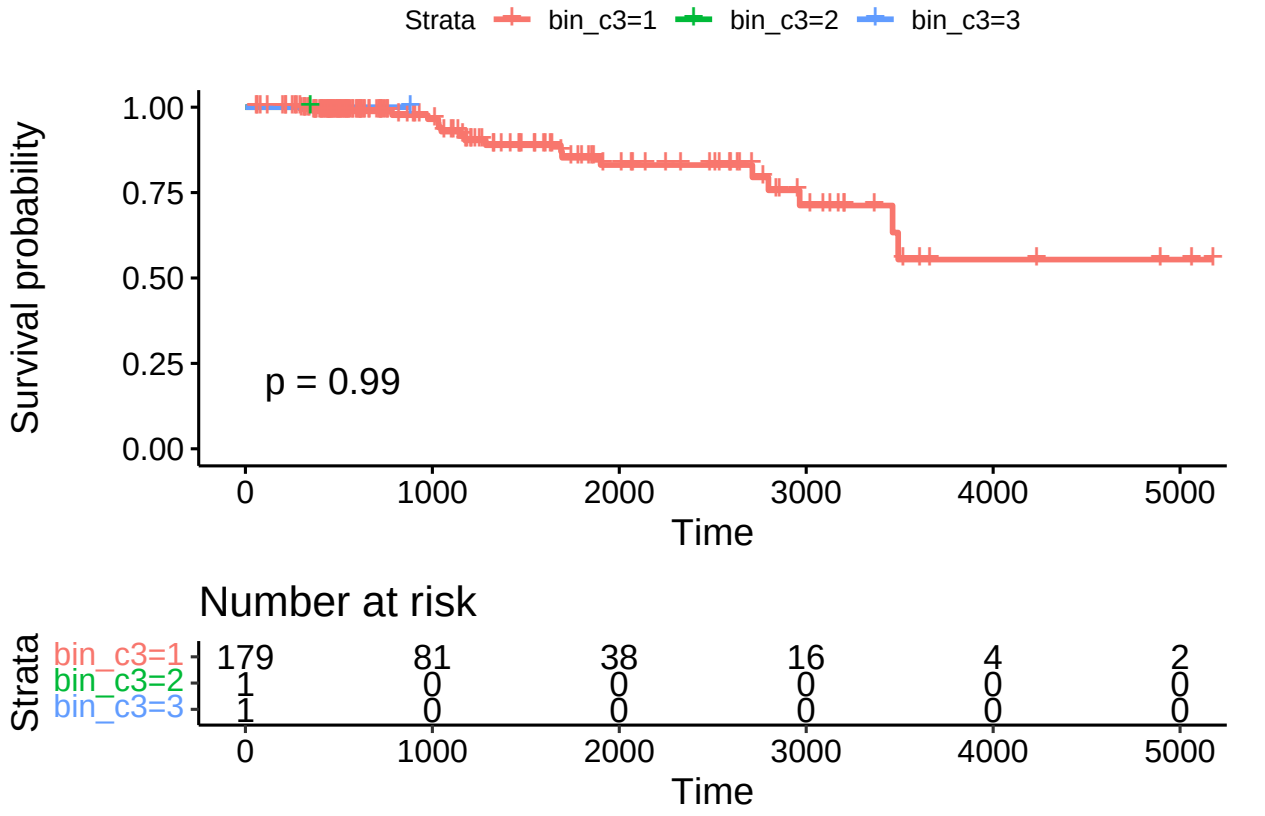


4.2.30. Binary distance measurement + Average Linkage

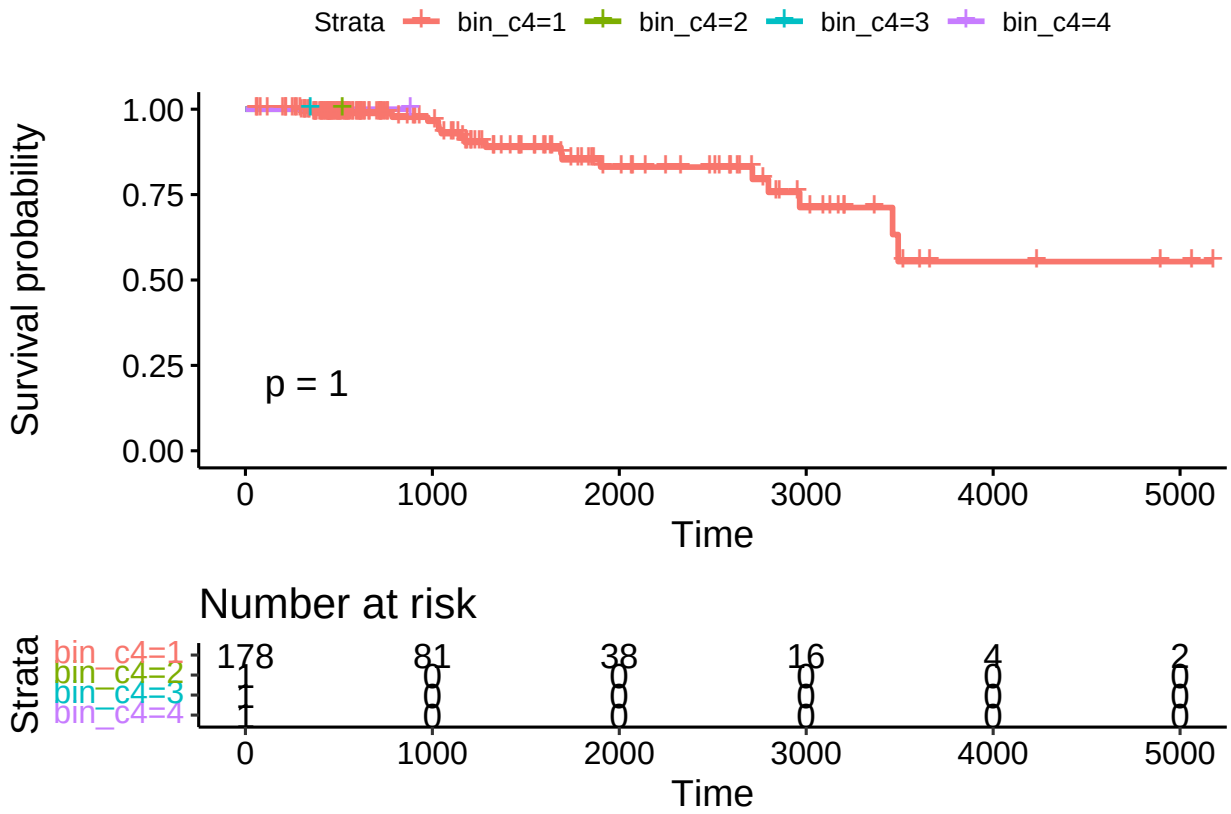
```
ggsurvplot(HC30,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS30,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

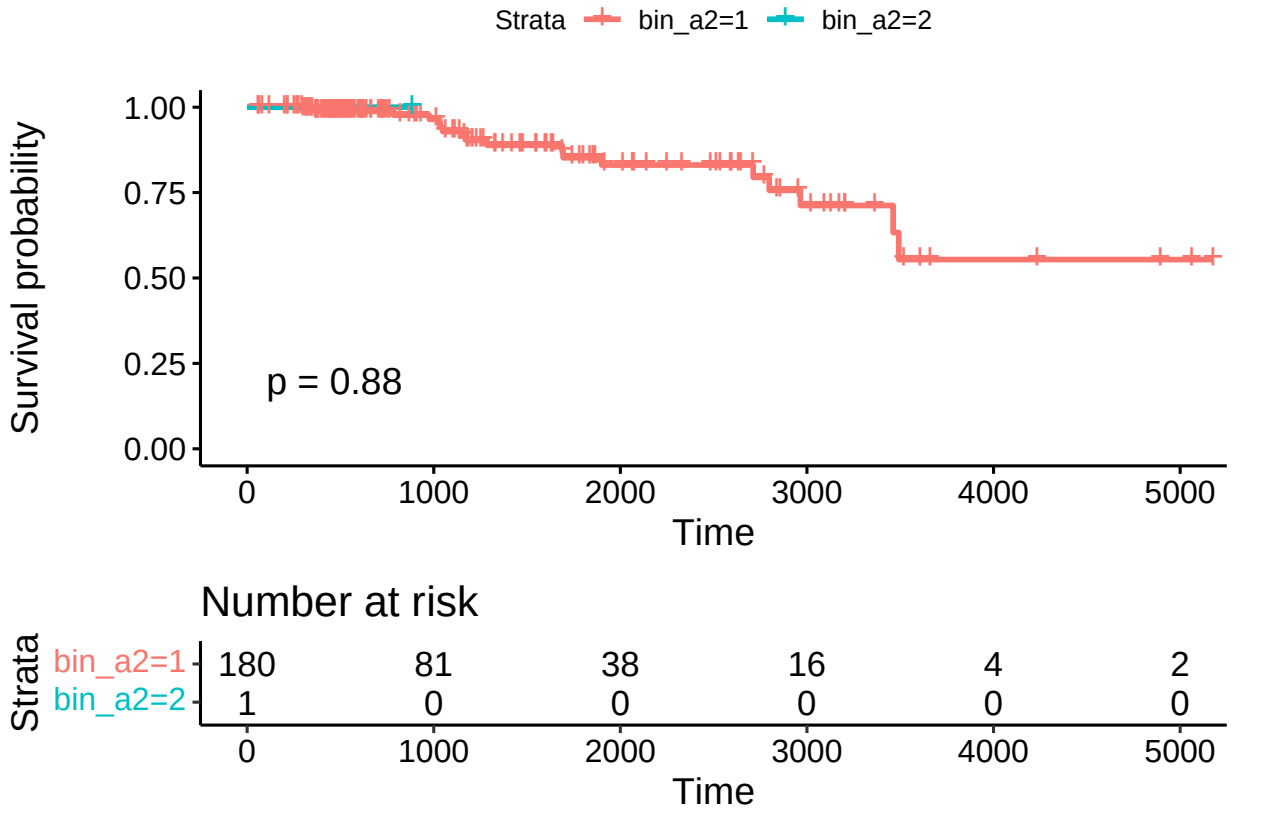



```
ggsurvplot(HW30,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

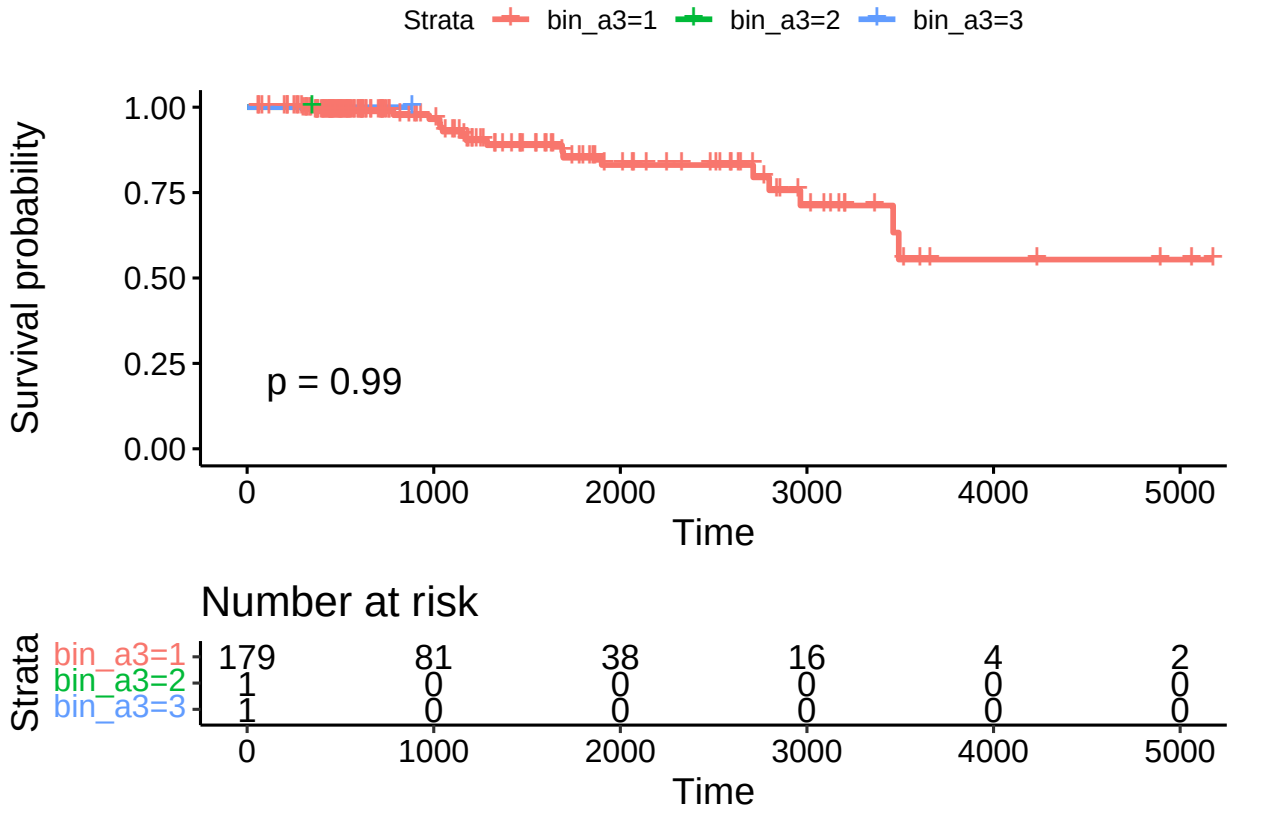


4.2.31. Binary distance measurement + Complete Linkage

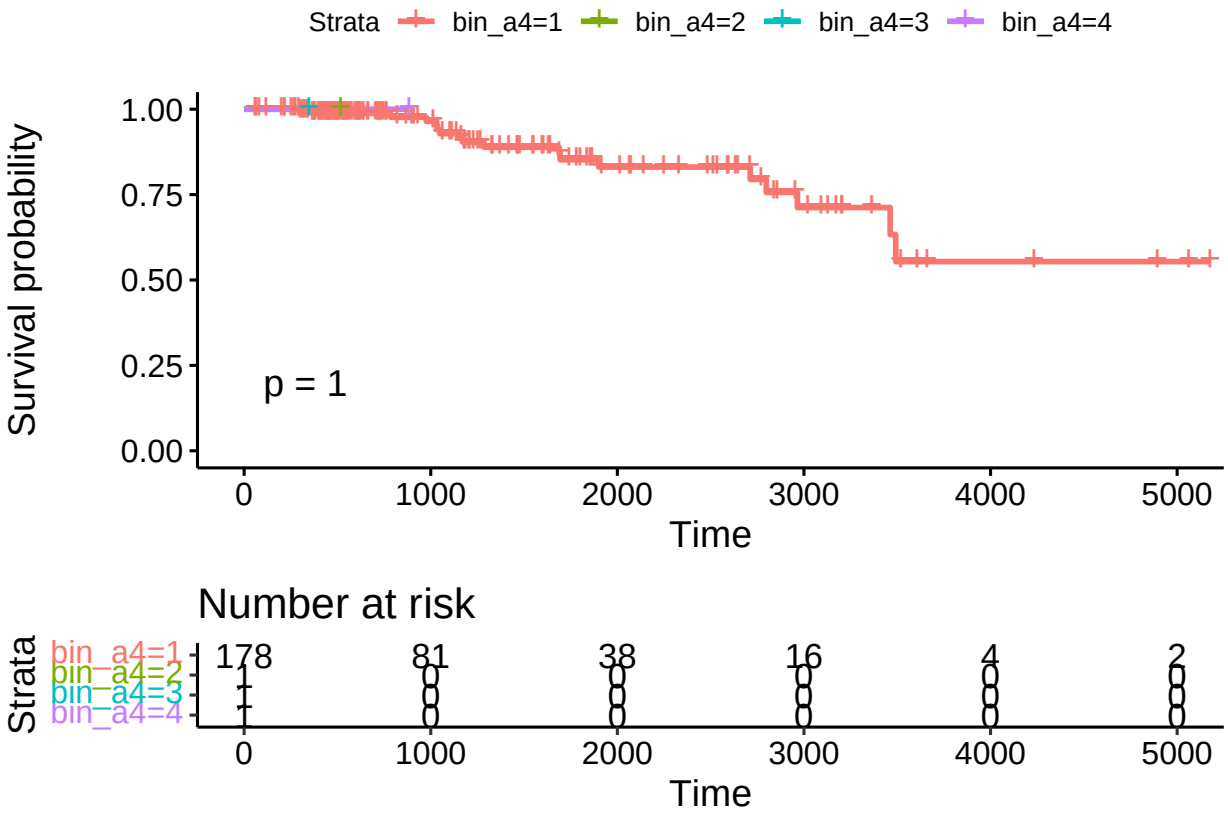
```
ggsurvplot(HC31,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS31,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

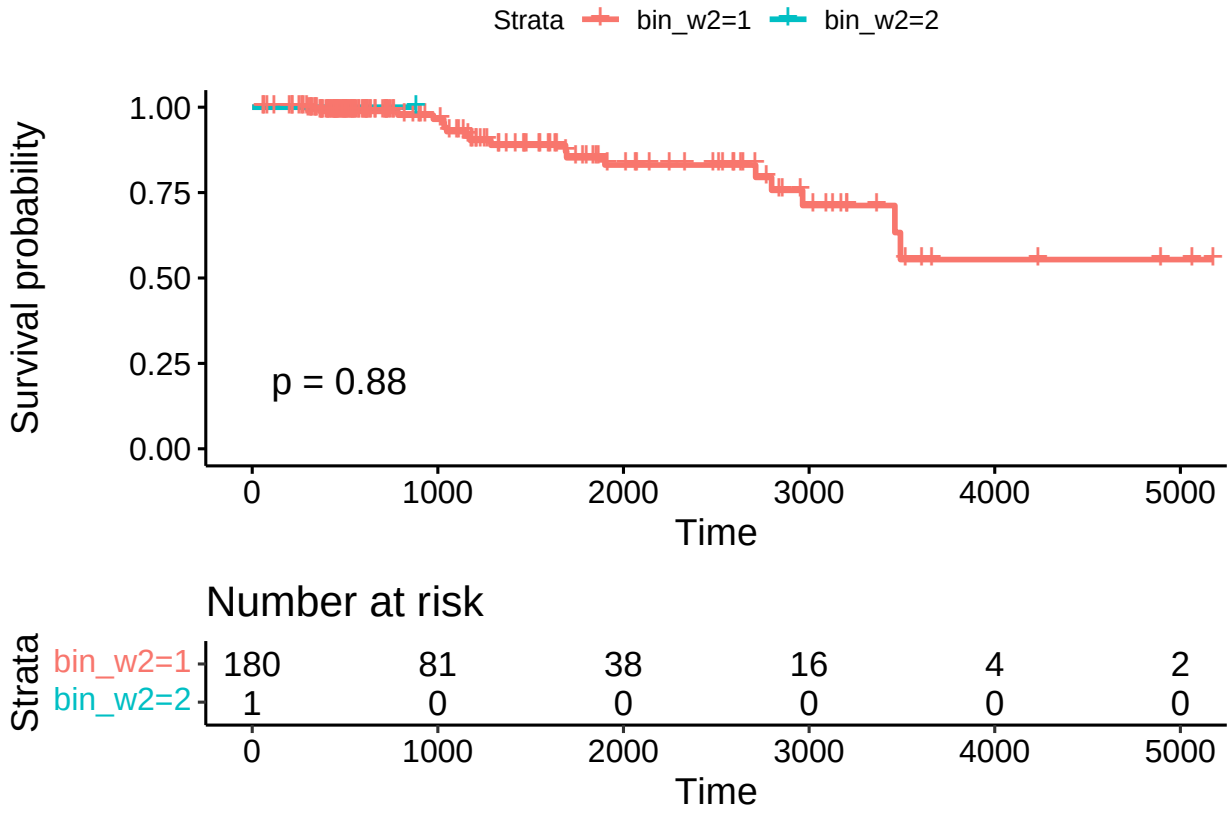



```
ggsurvplot(HW31,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

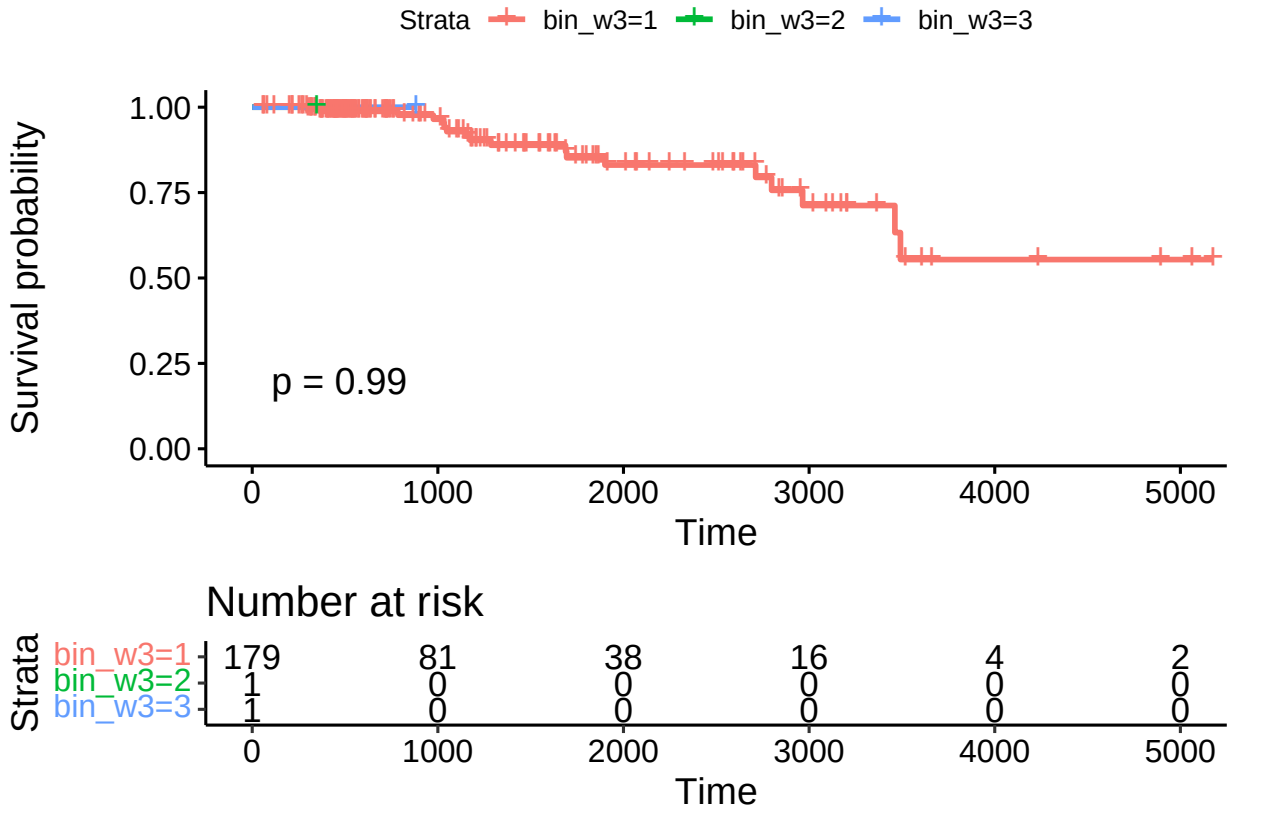


4.2.32. Binary distance measurement + Ward Linkage

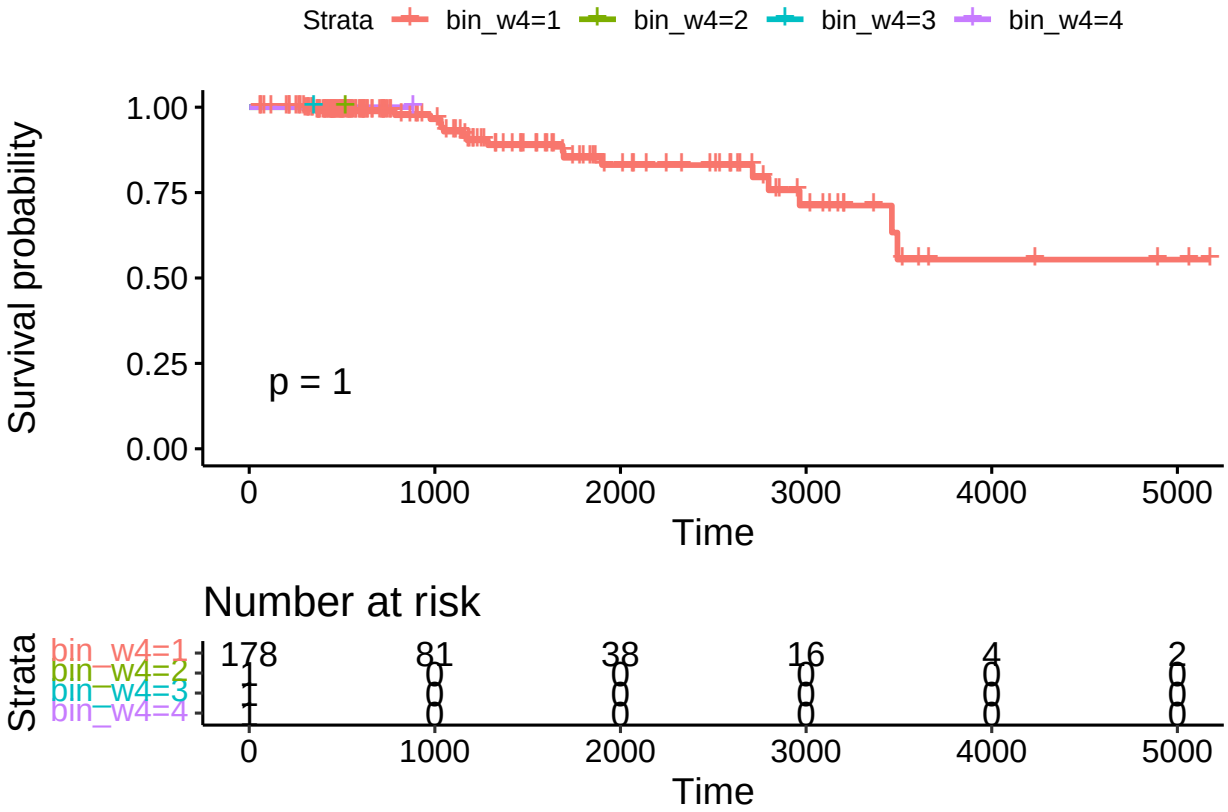
```
ggsurvplot(HC32,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS32,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

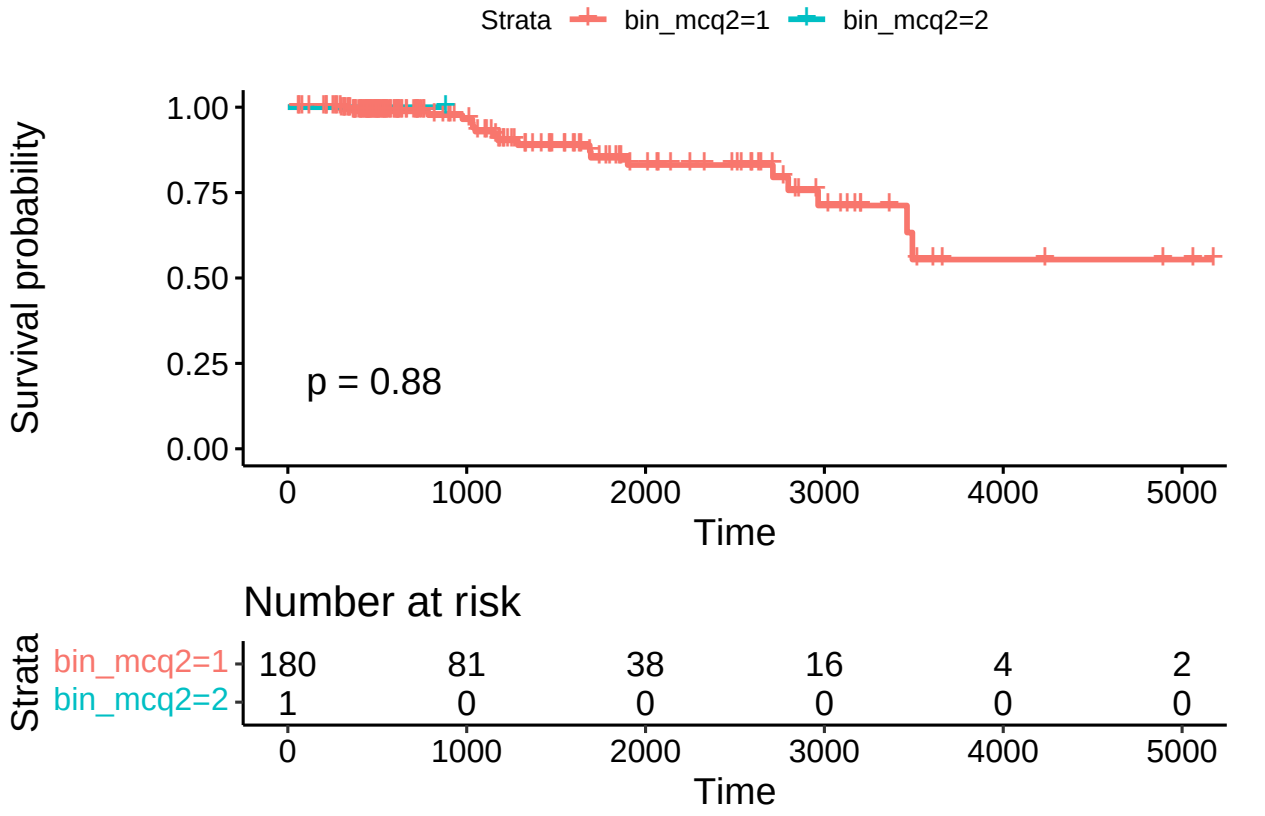



```
ggsurvplot(HW32,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

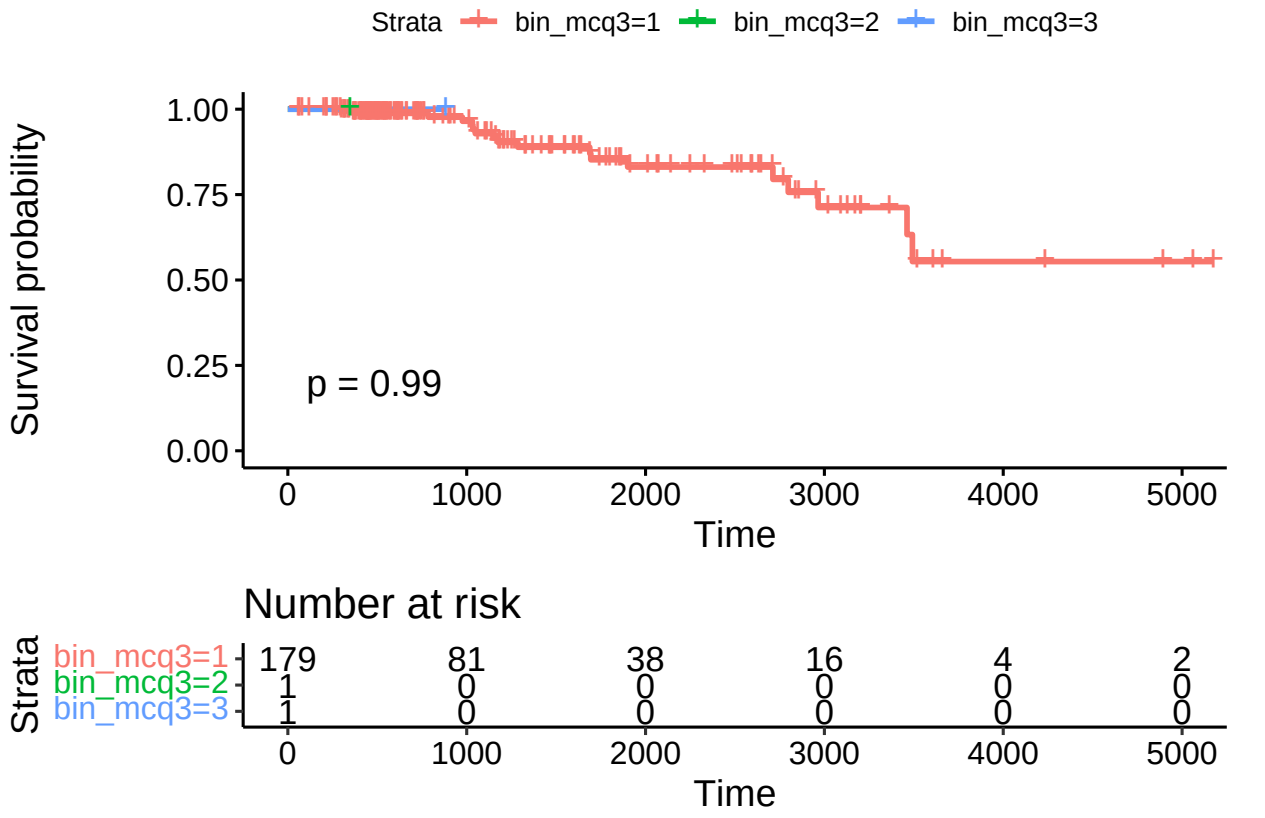


4.2.33. Binary distance measurement + Mcquitty Linkage

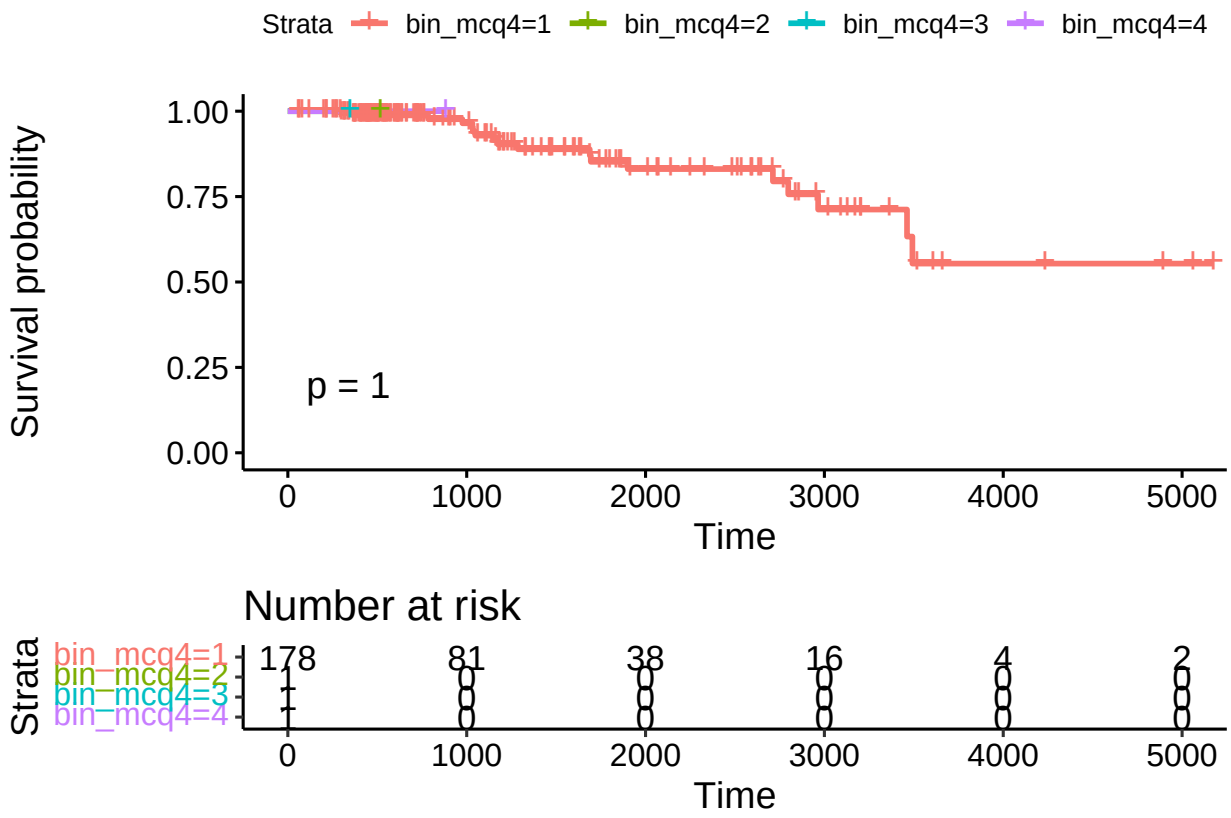
```
ggsurvplot(HC33,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS33,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

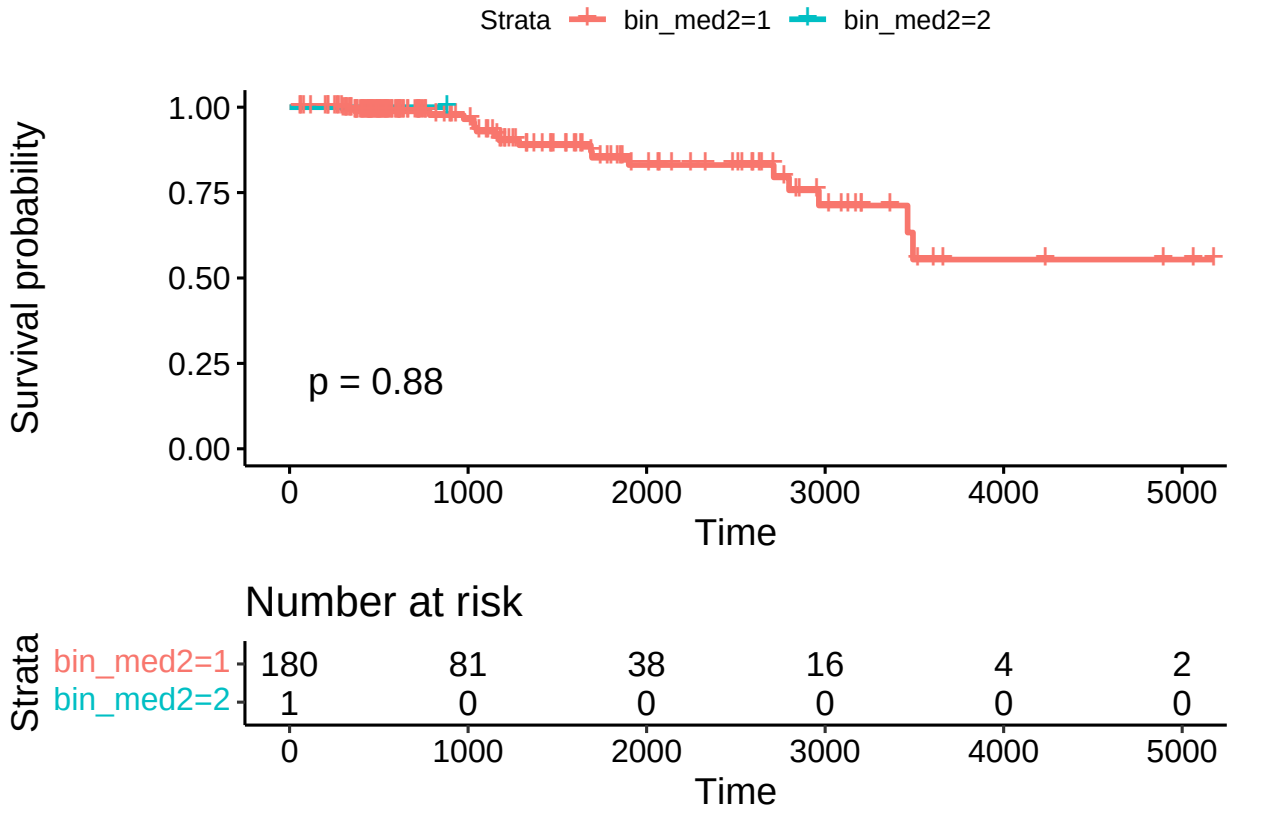



```
ggsurvplot(HW33,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

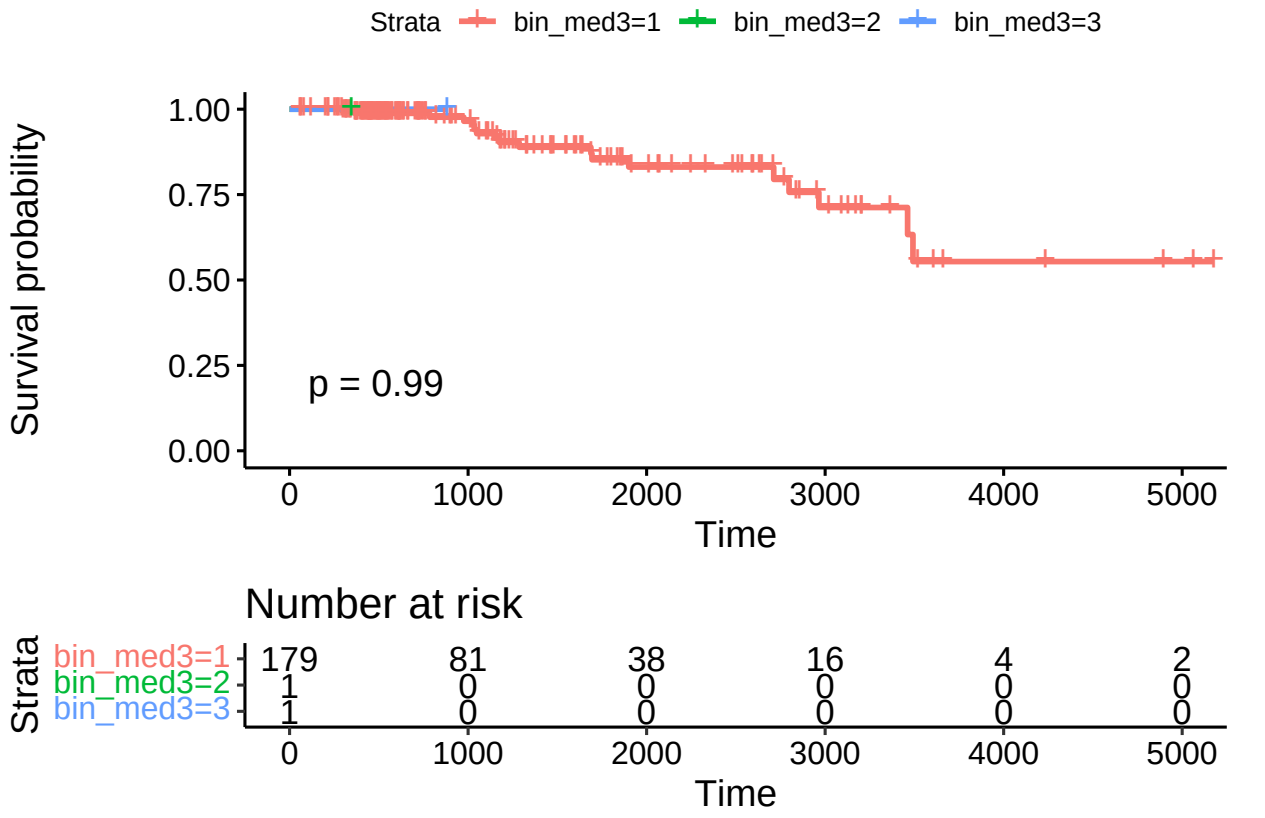


4.2.34. Binary distance measurement + Median Linkage

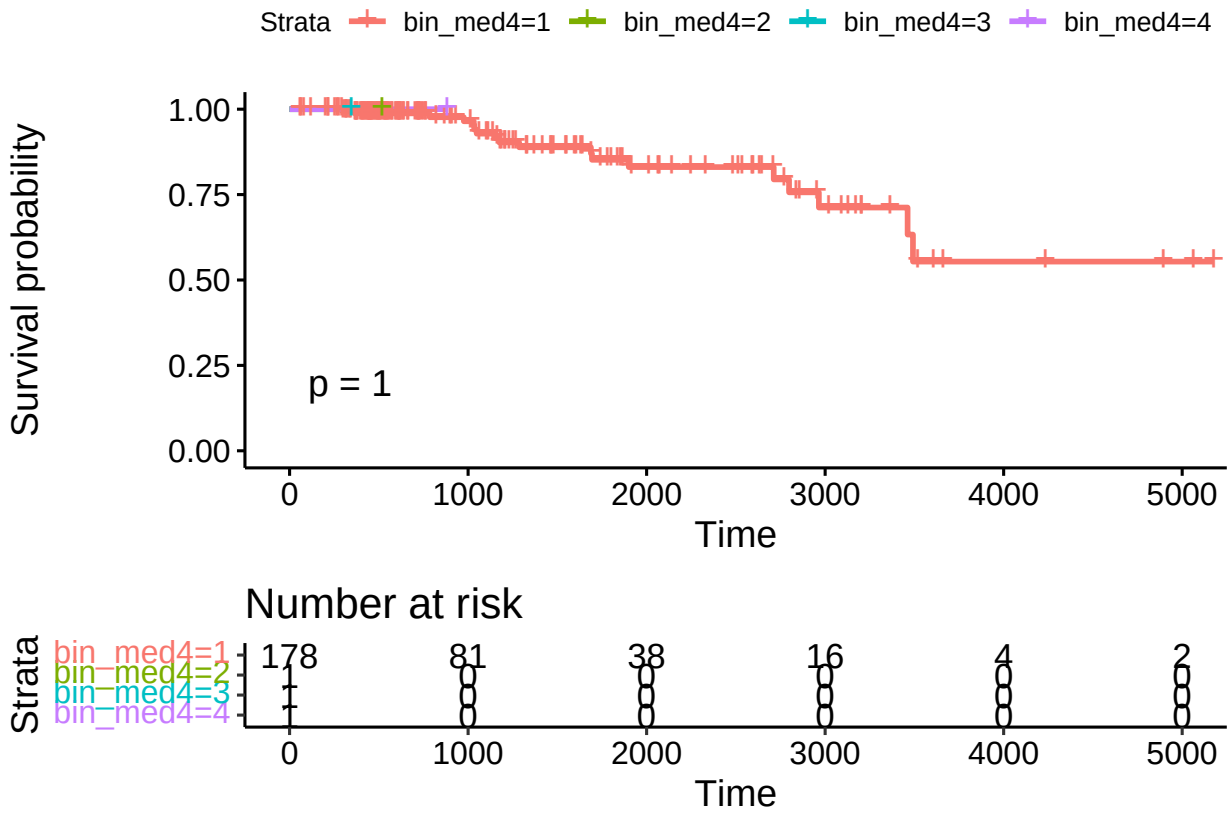
```
ggsurvplot(HC34,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS34,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

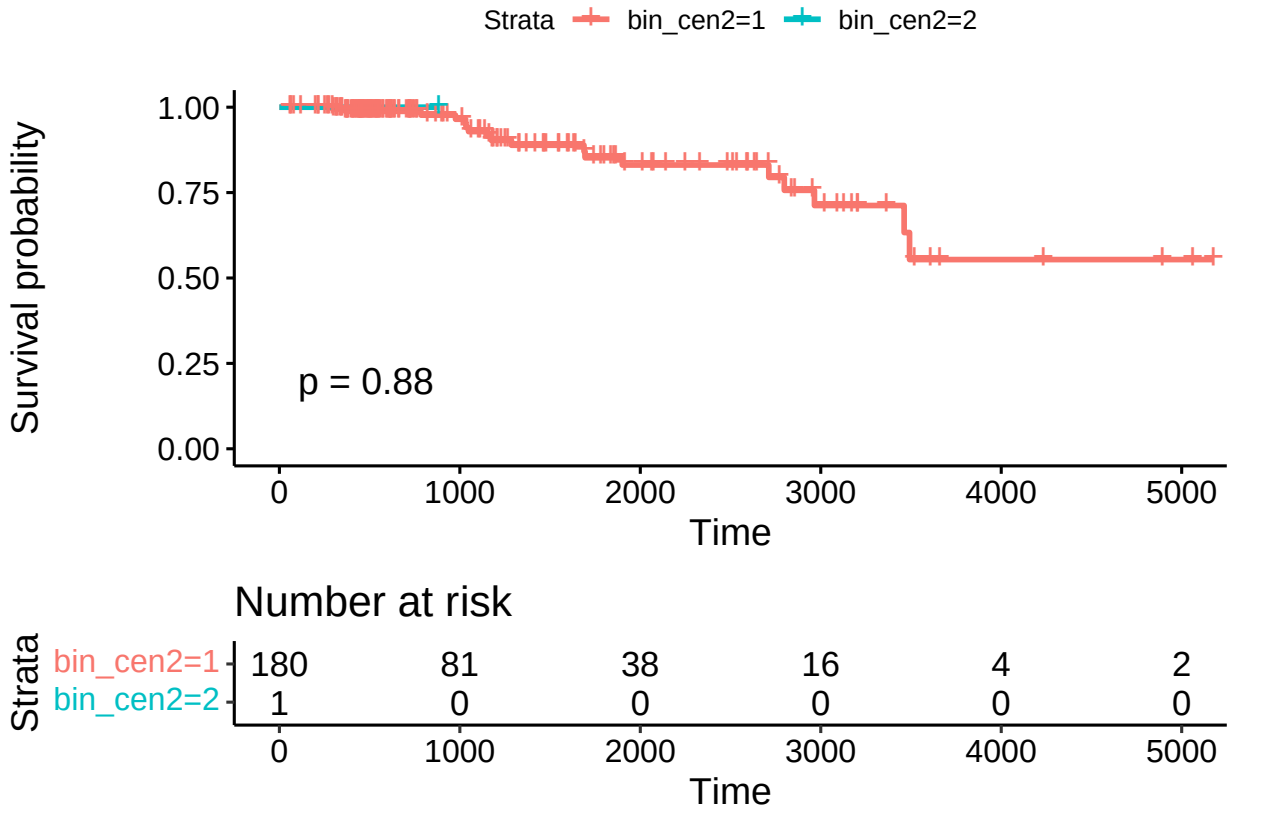



```
ggsurvplot(HW34,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

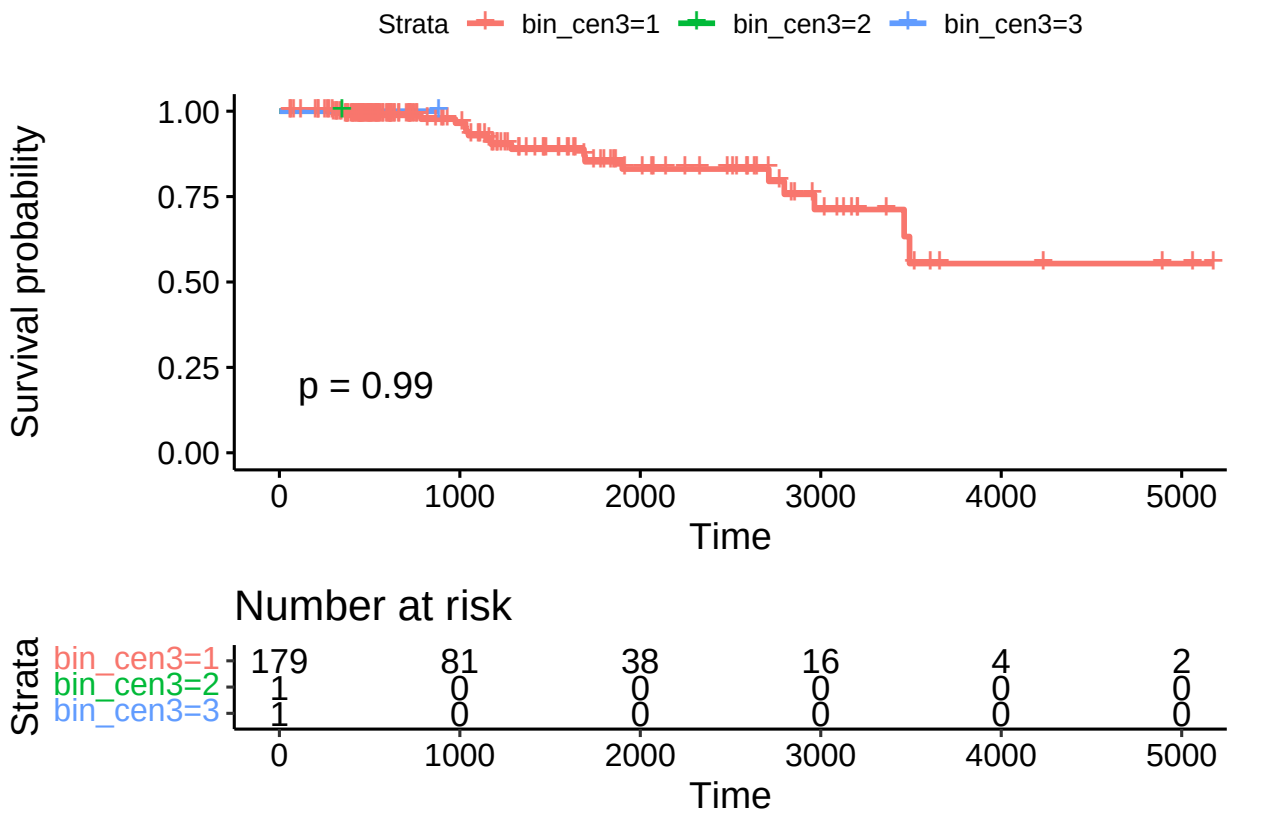


4.2.35. Binary distance measurement + Centroid Linkage

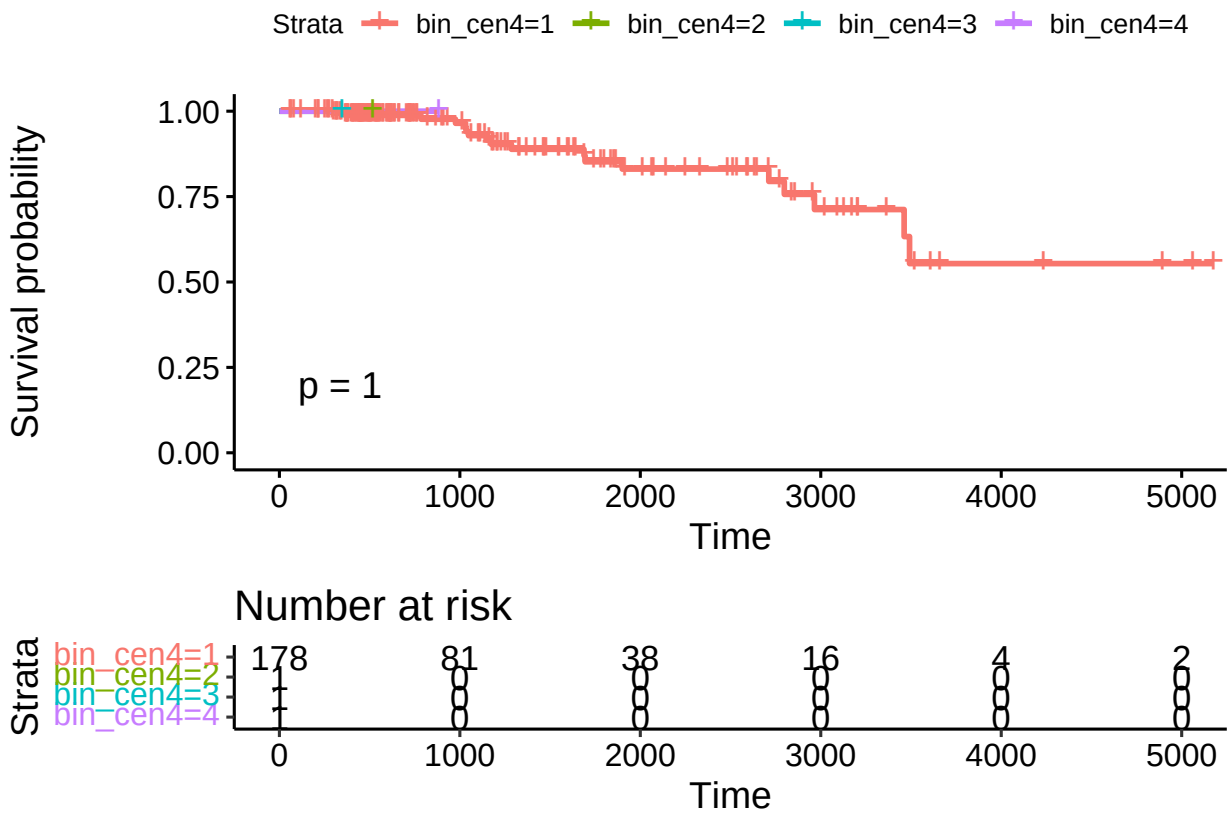
```
ggsurvplot(HC35,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS35,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

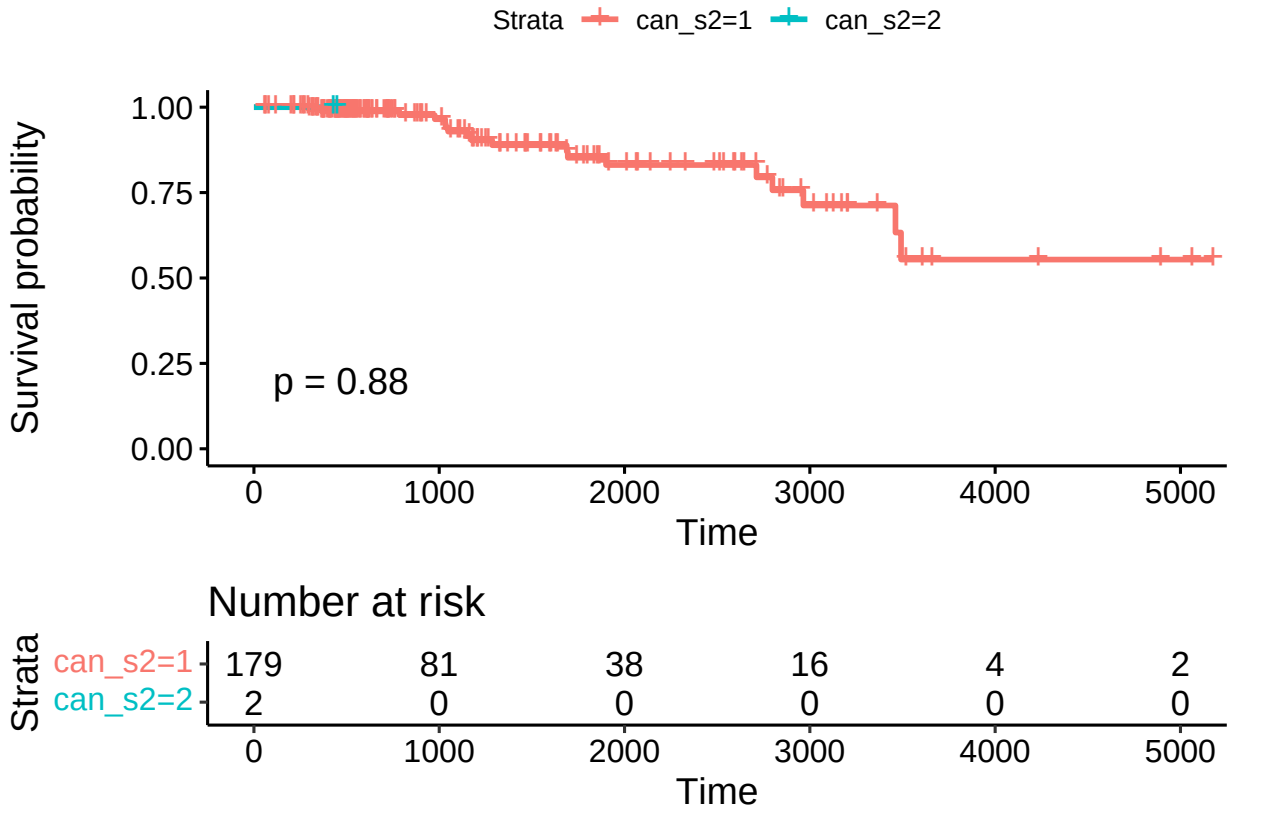



```
ggsurvplot(HW35,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

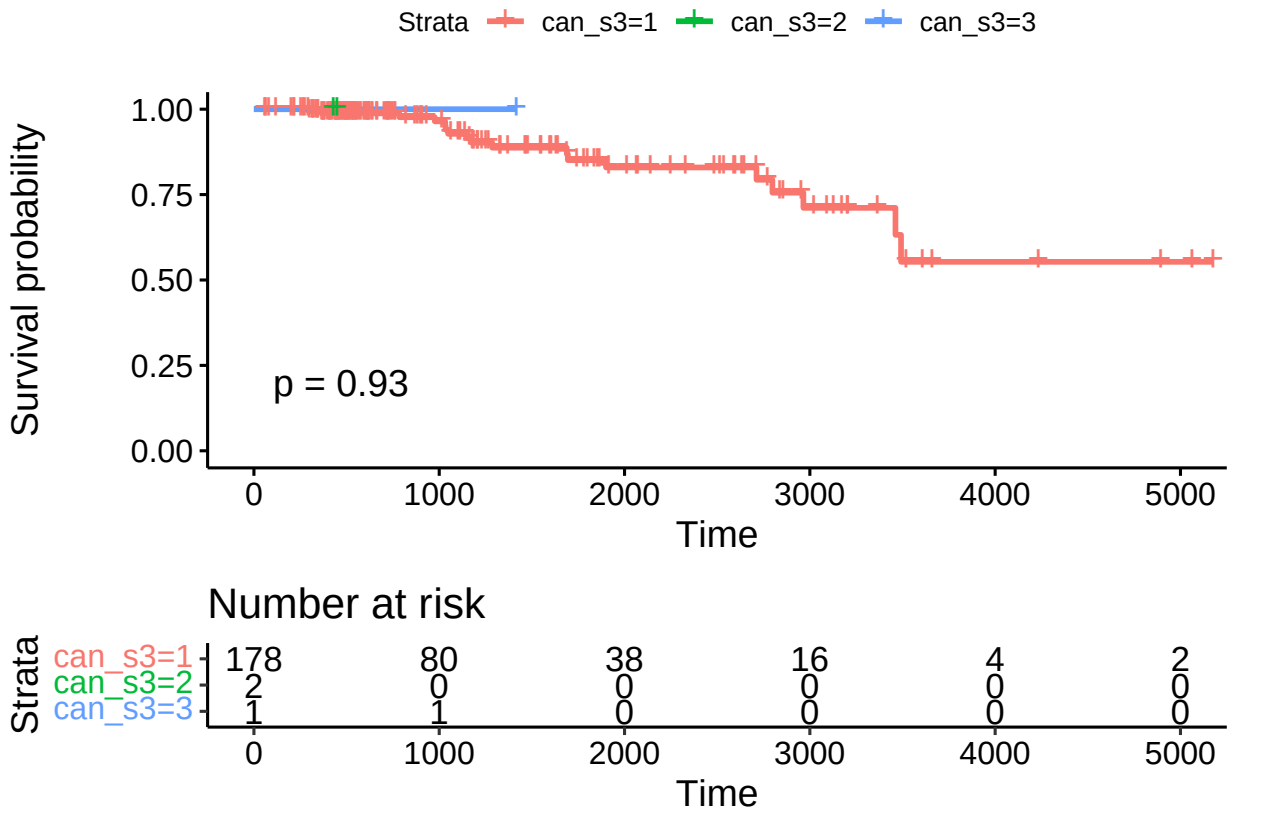


4.2.36. Canberra distance measurement + Single Linkage

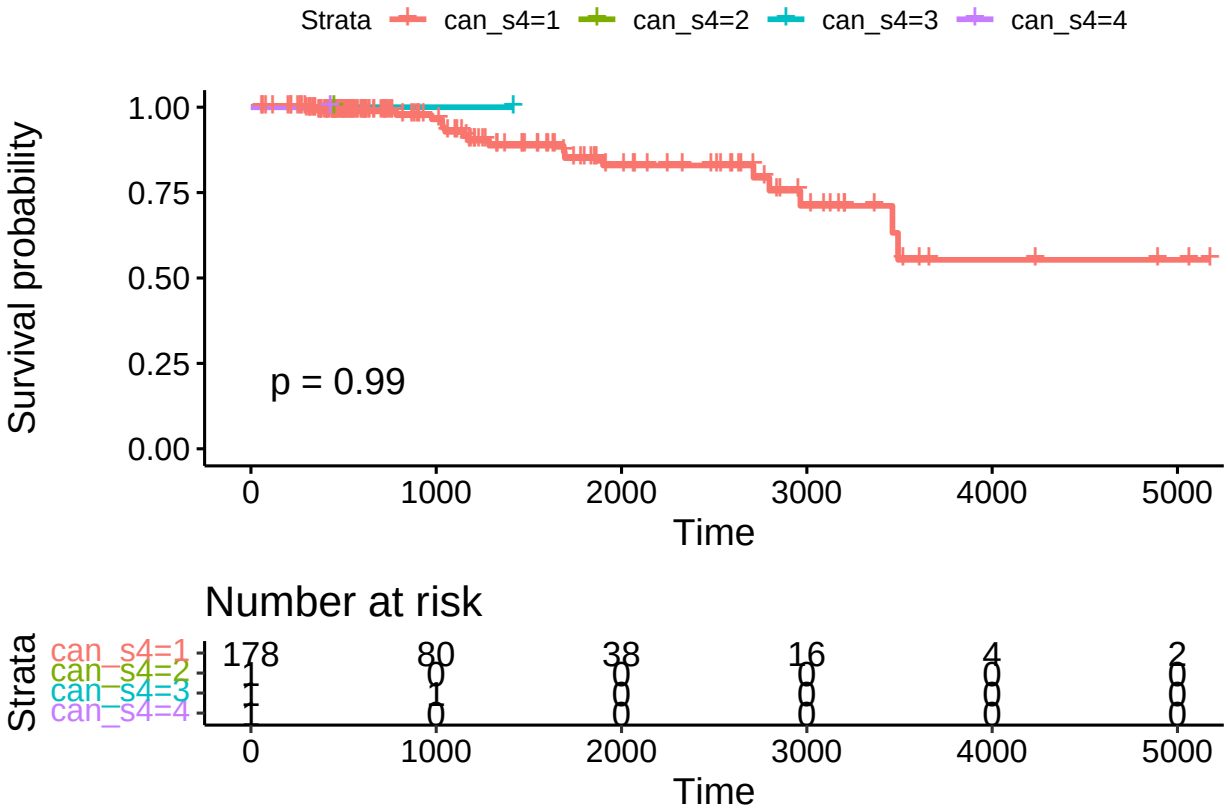
```
ggsurvplot(HC36,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS36,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

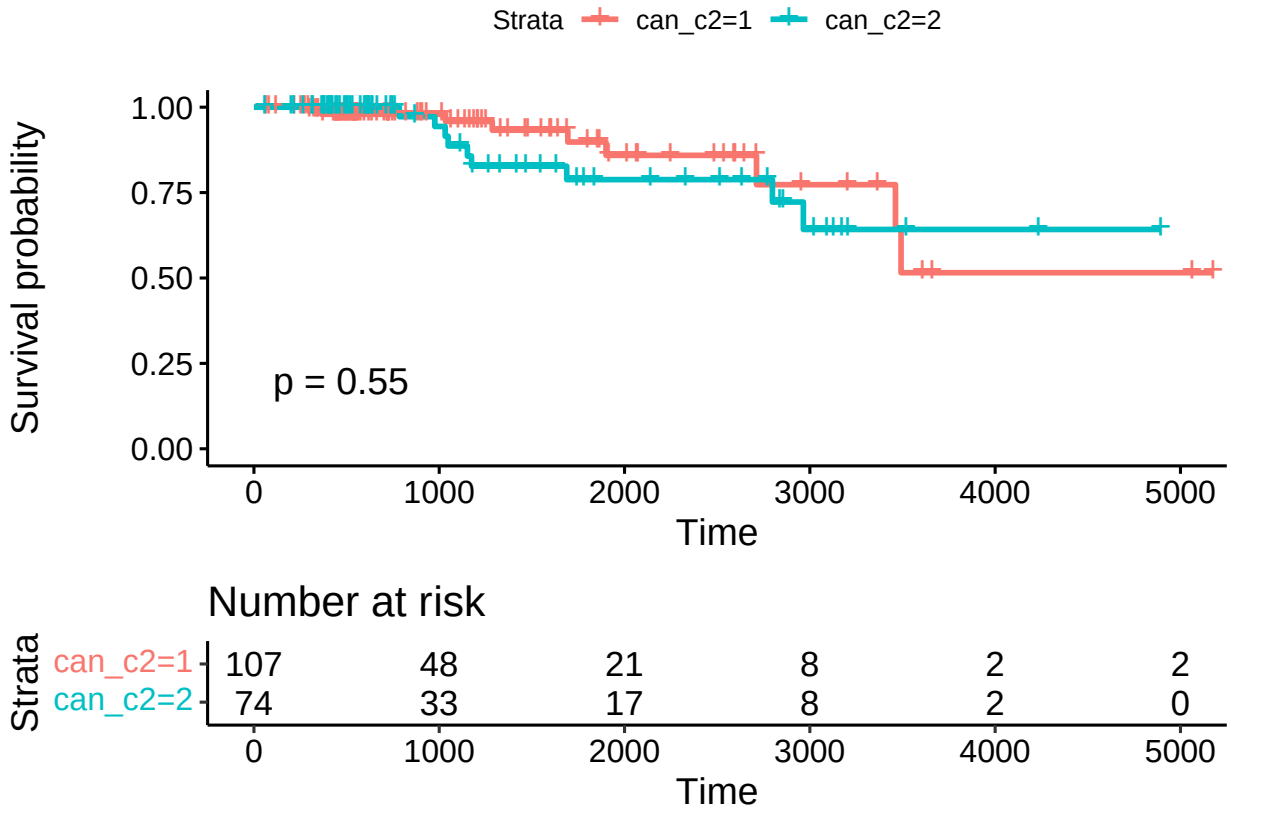



```
ggsurvplot(HW36,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

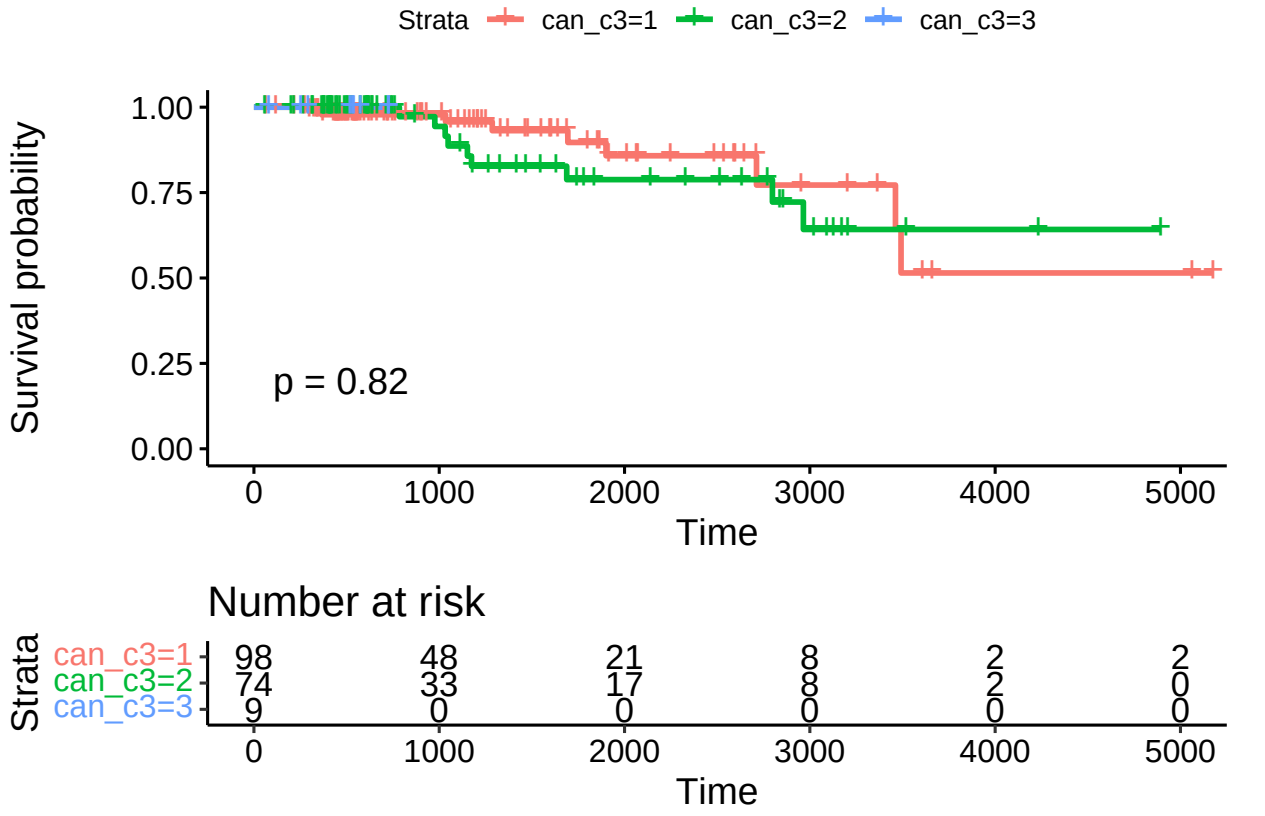


4.2.37. Canberra distance measurement + Average Linkage

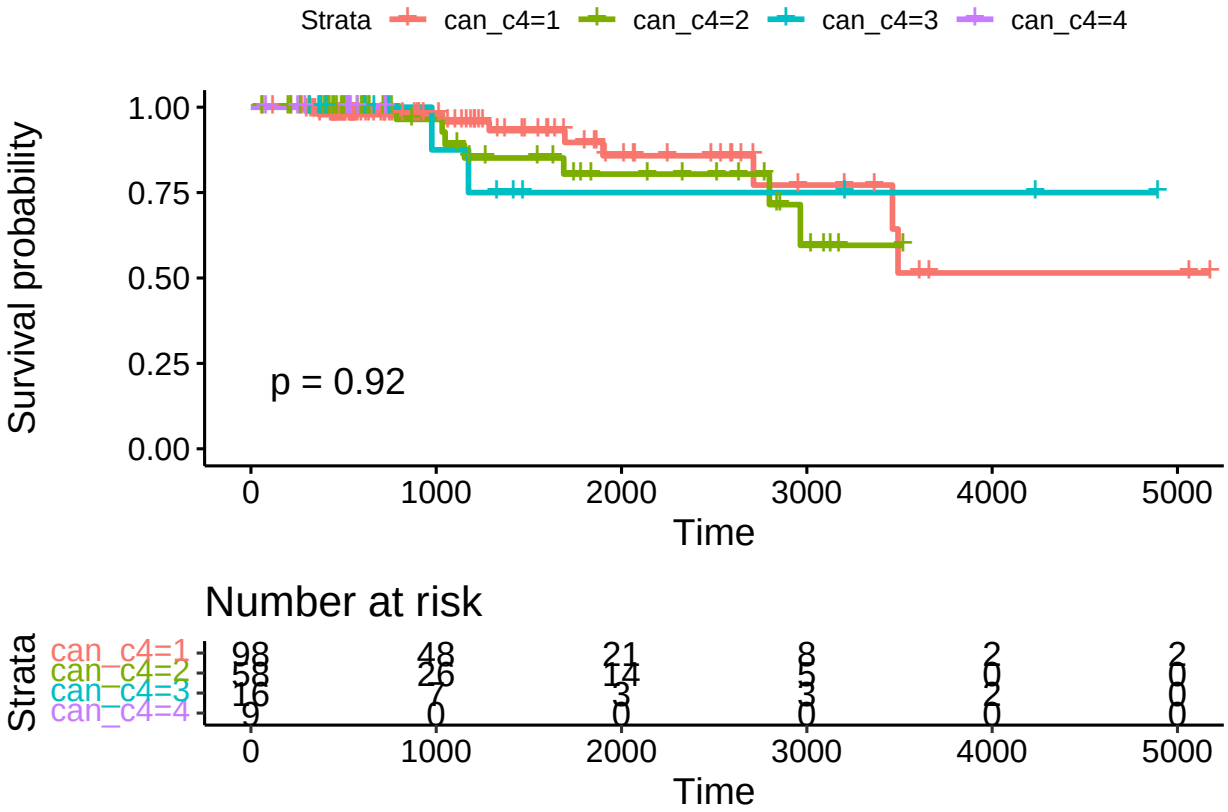
```
ggsurvplot(HC37,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS37,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

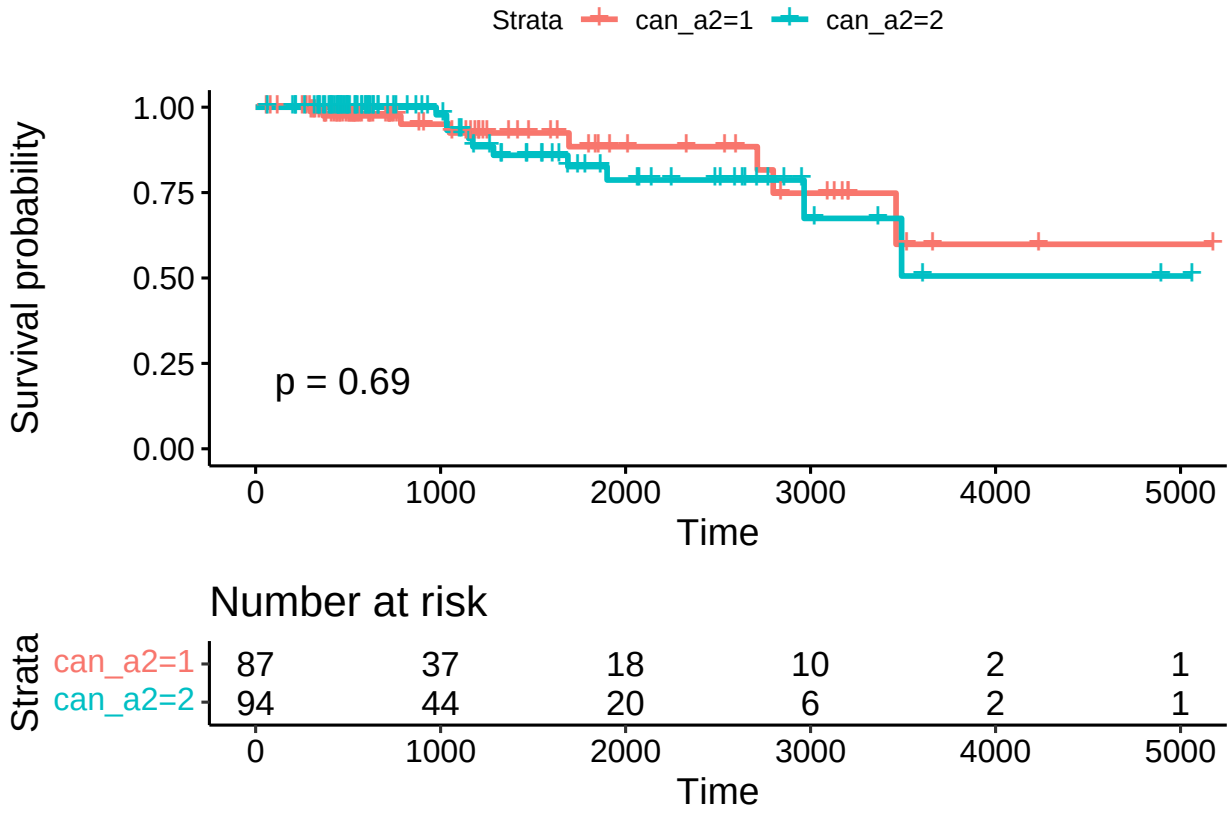



```
ggsurvplot(HW37,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

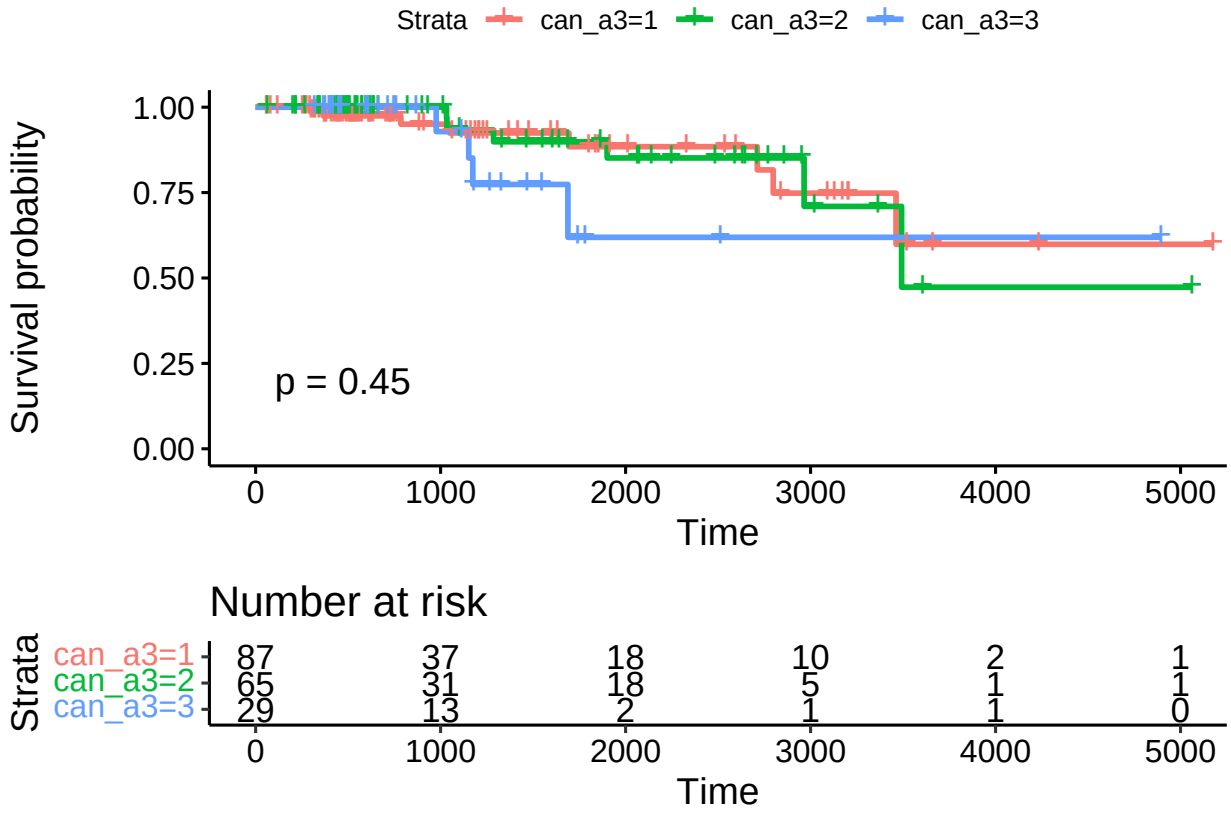


4.2.38. Canberra distance measurement + Complete Linkage

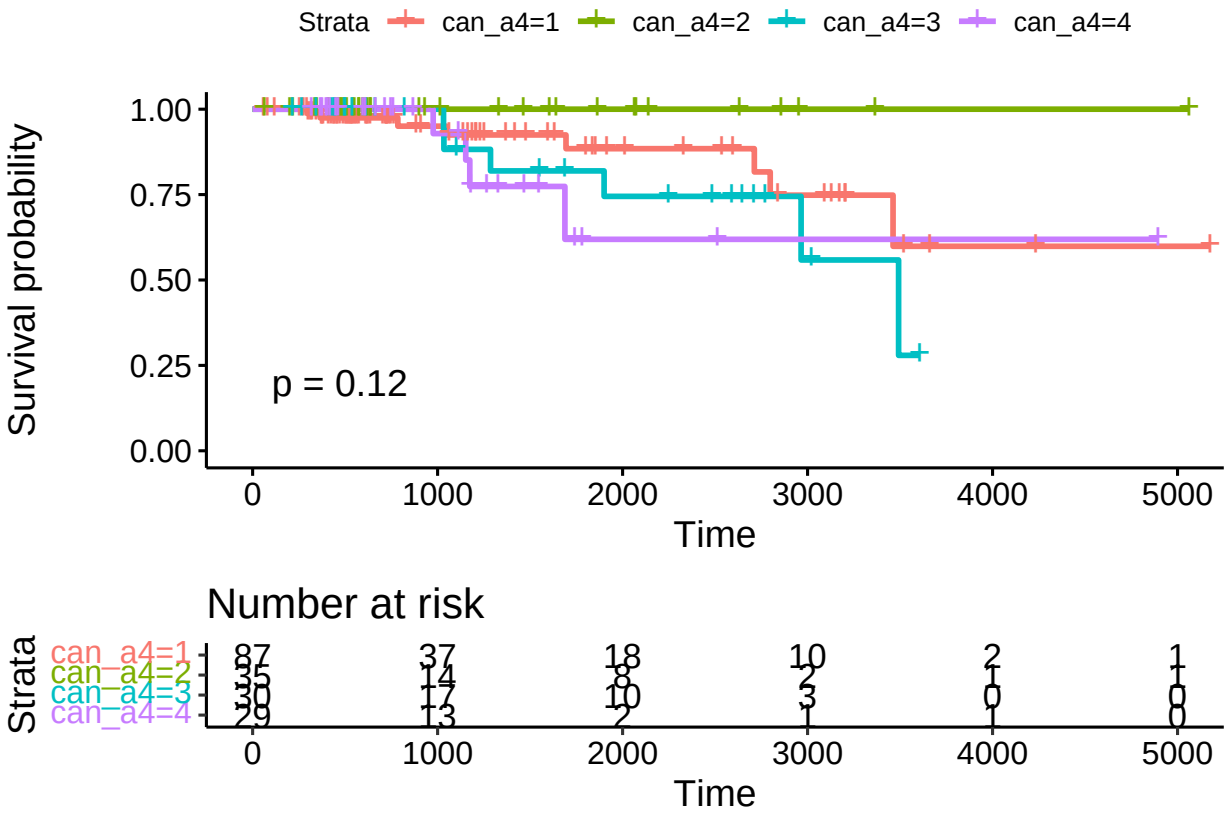
```
ggsurvplot(HC38,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS38,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

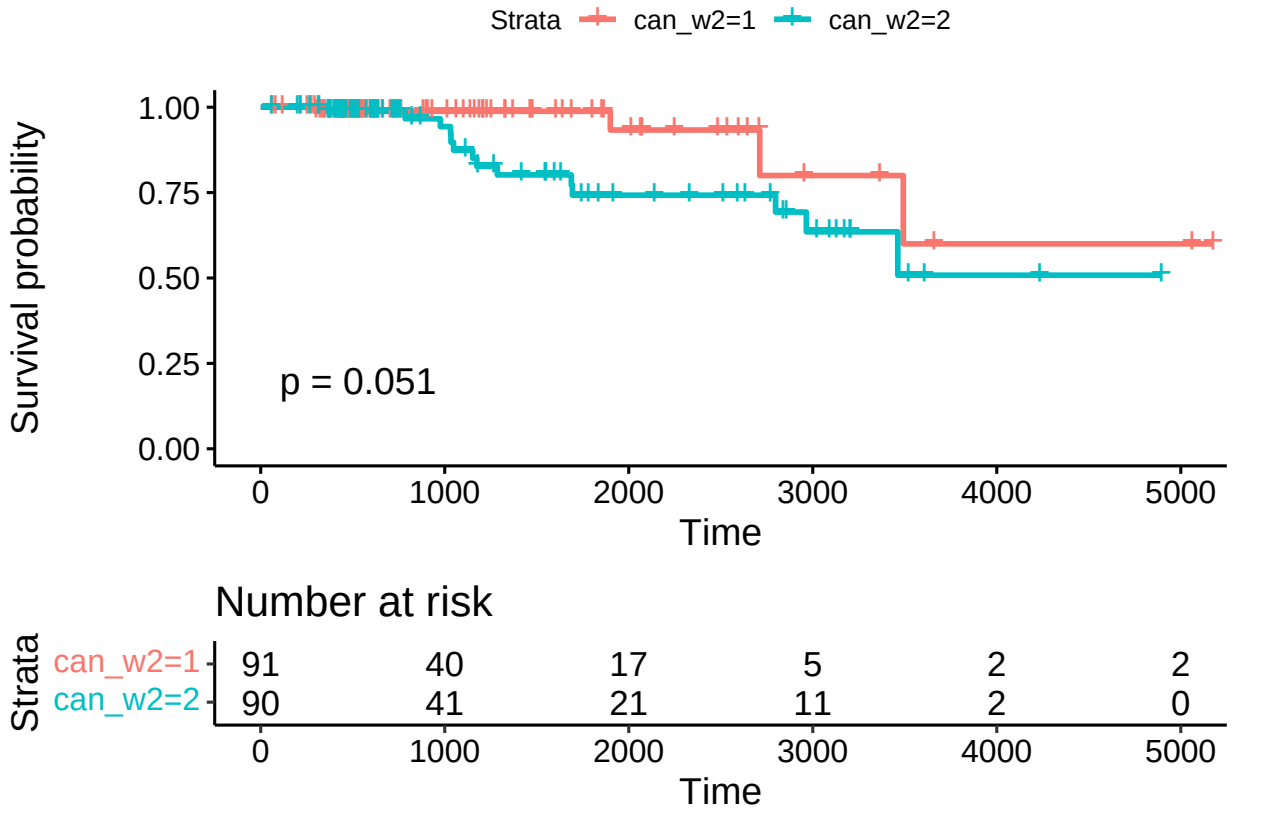



```
ggsurvplot(HW38,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

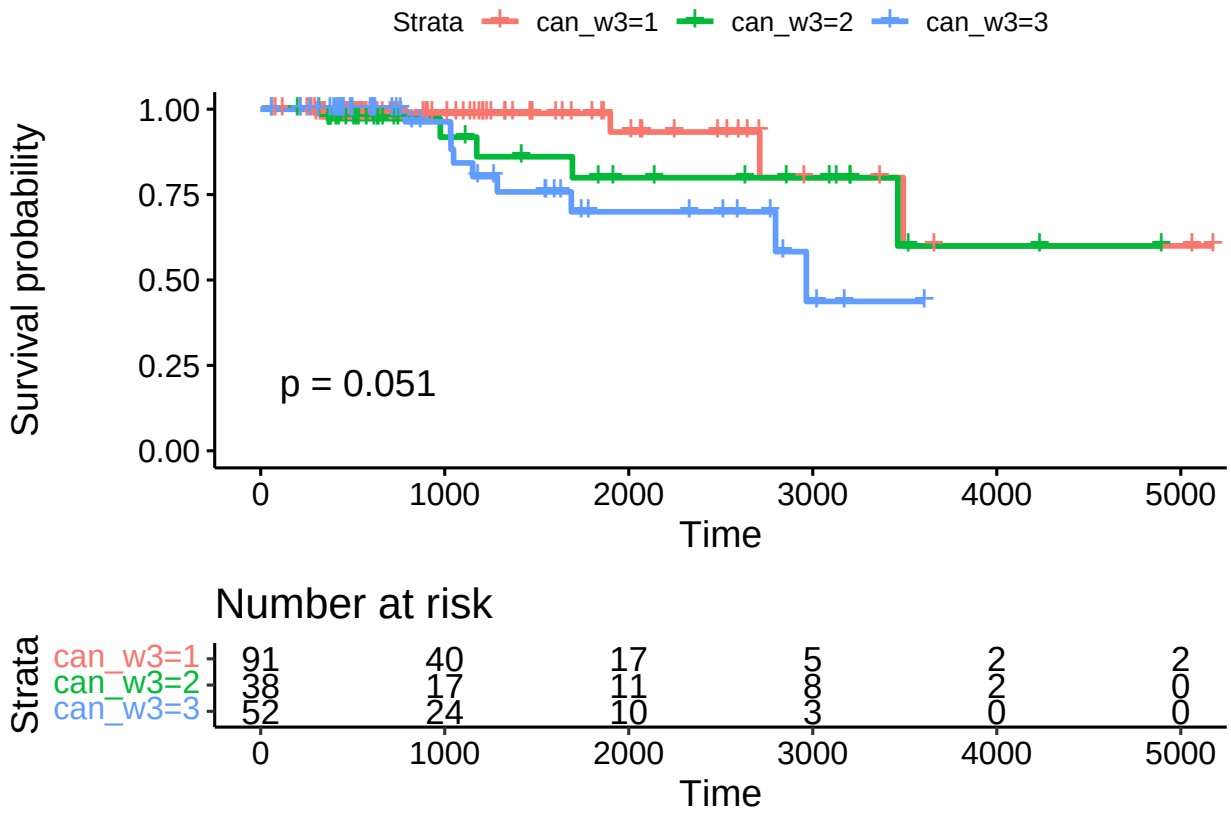


4.2.39. Canberra distance measurement + Ward Linkage

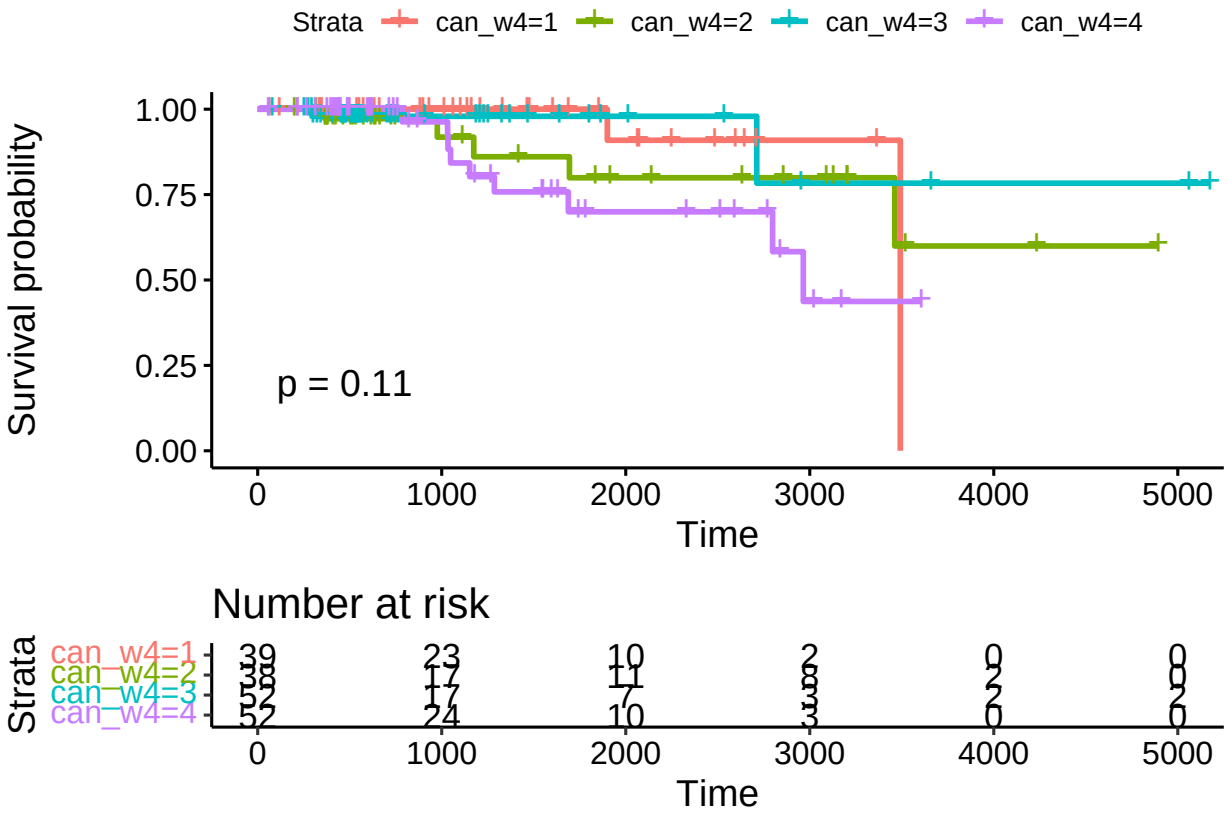
```
ggsurvplot(HC39,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS39,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

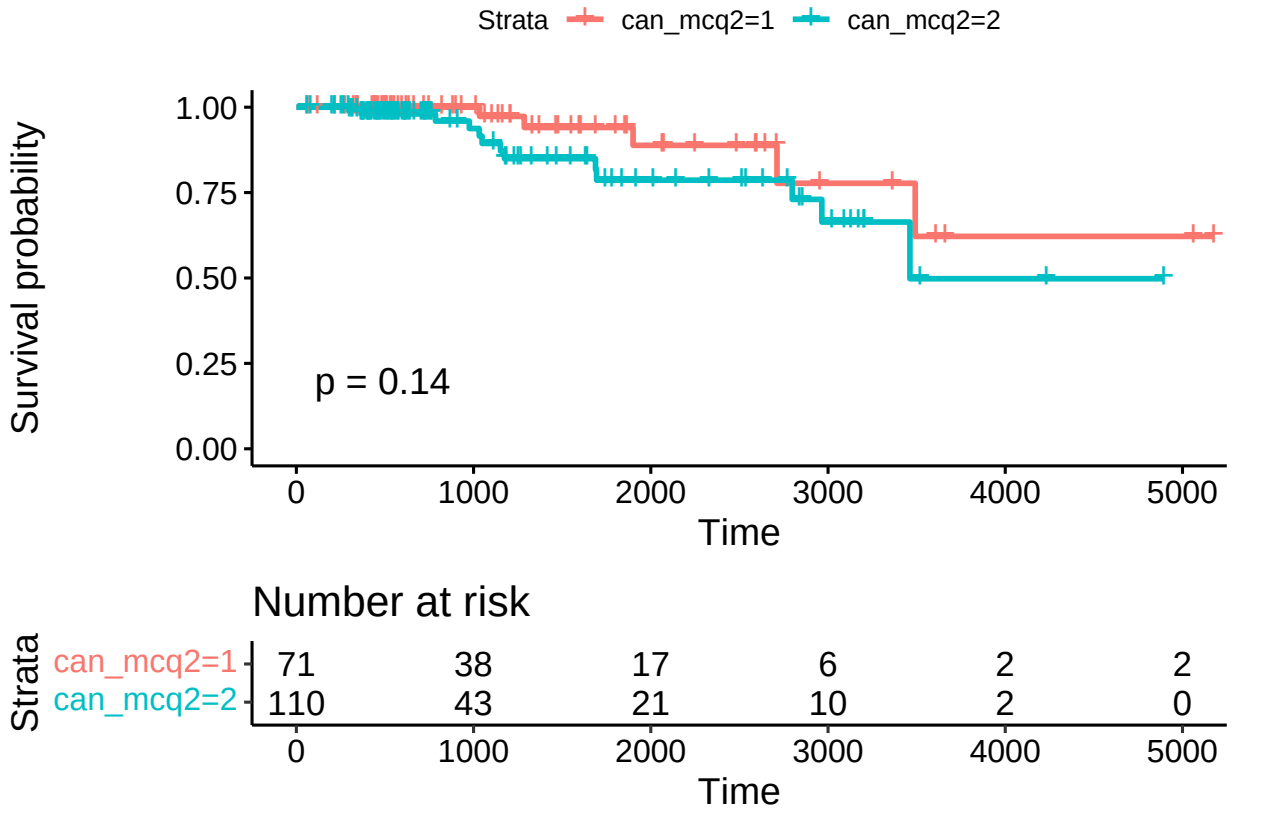



```
ggsurvplot(HW39,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

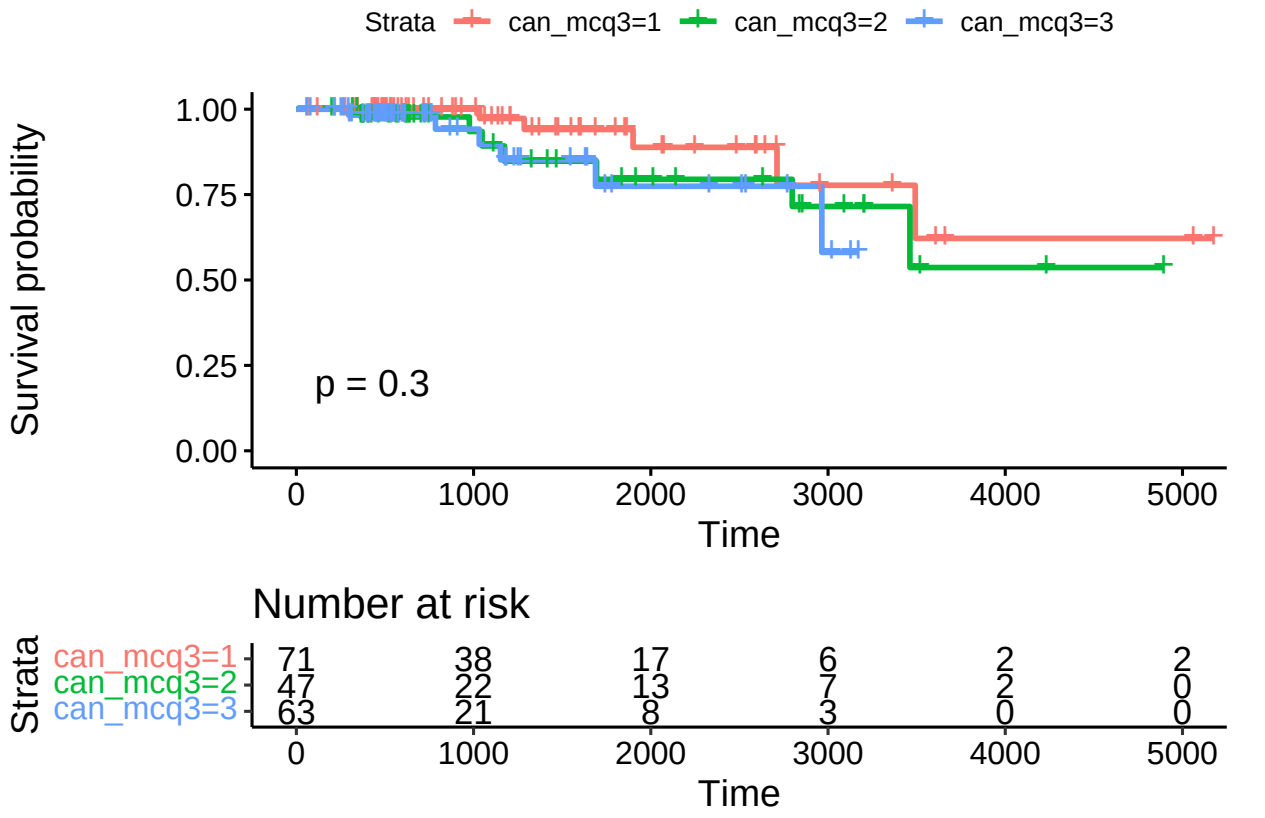


4.2.40. Canberra distance measurement + Mcquitty Linkage

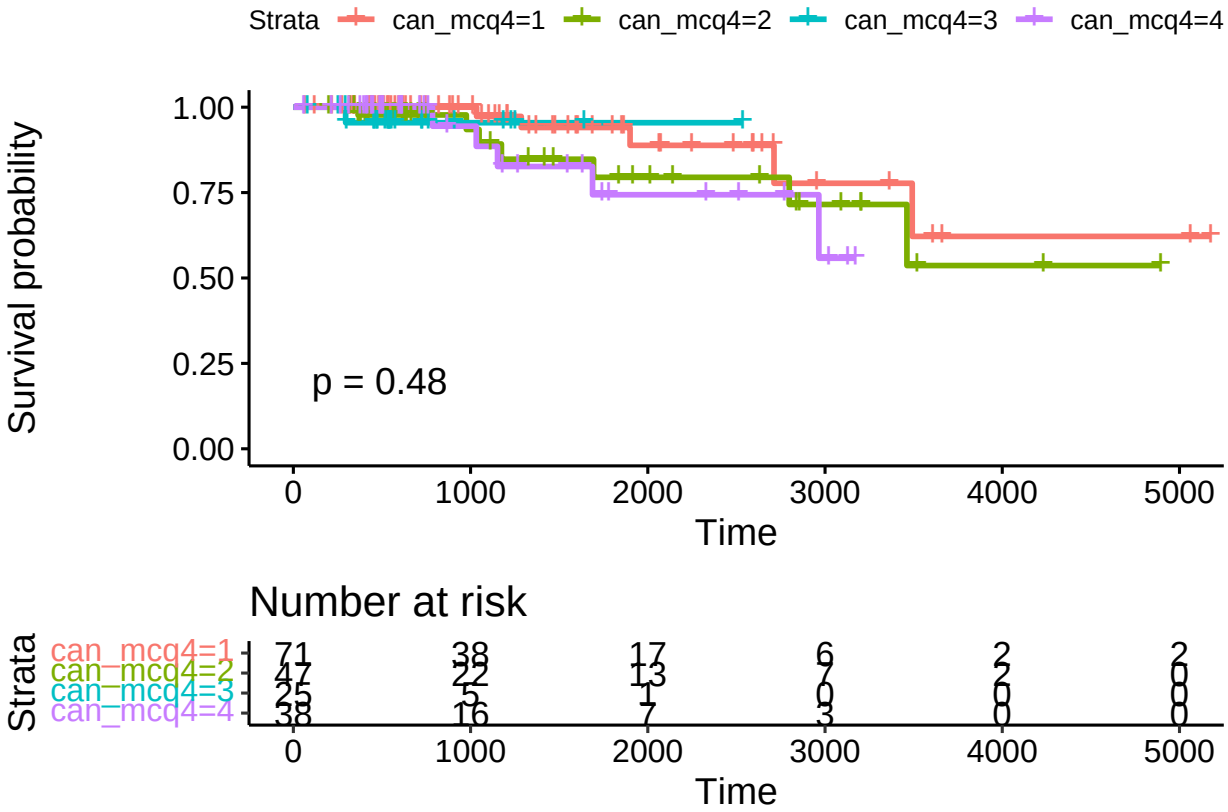
```
ggsurvplot(HC40,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS40,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

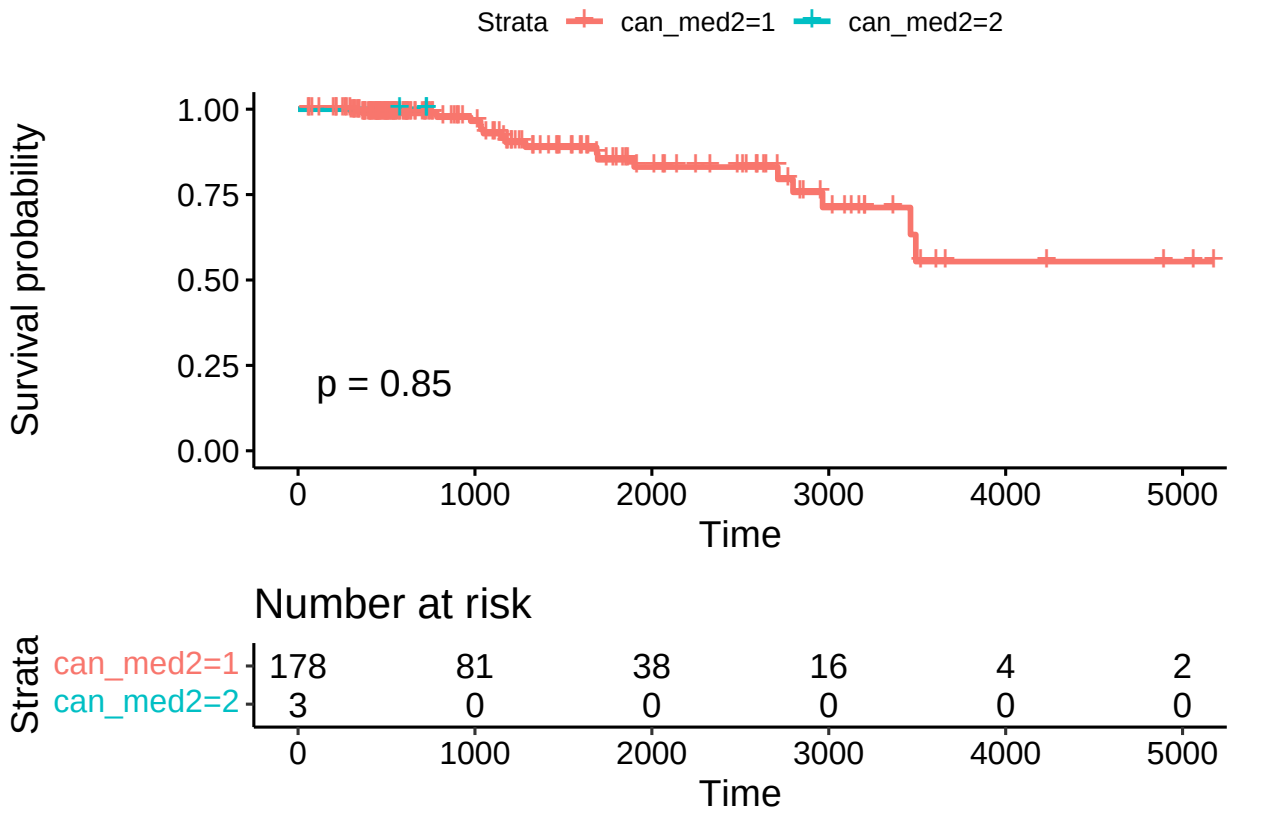



```
ggsurvplot(HW40,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

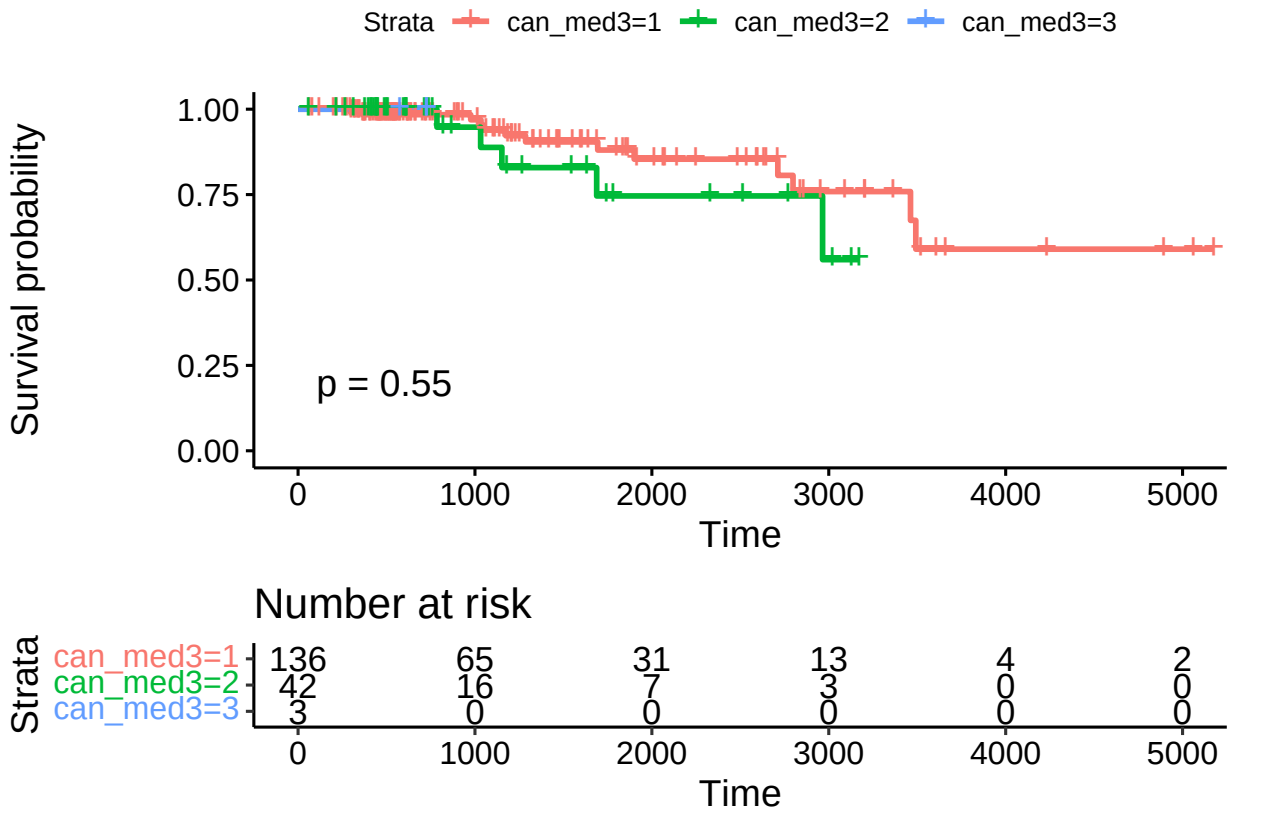


4.2.41. Canberra distance measurement + Median Linkage

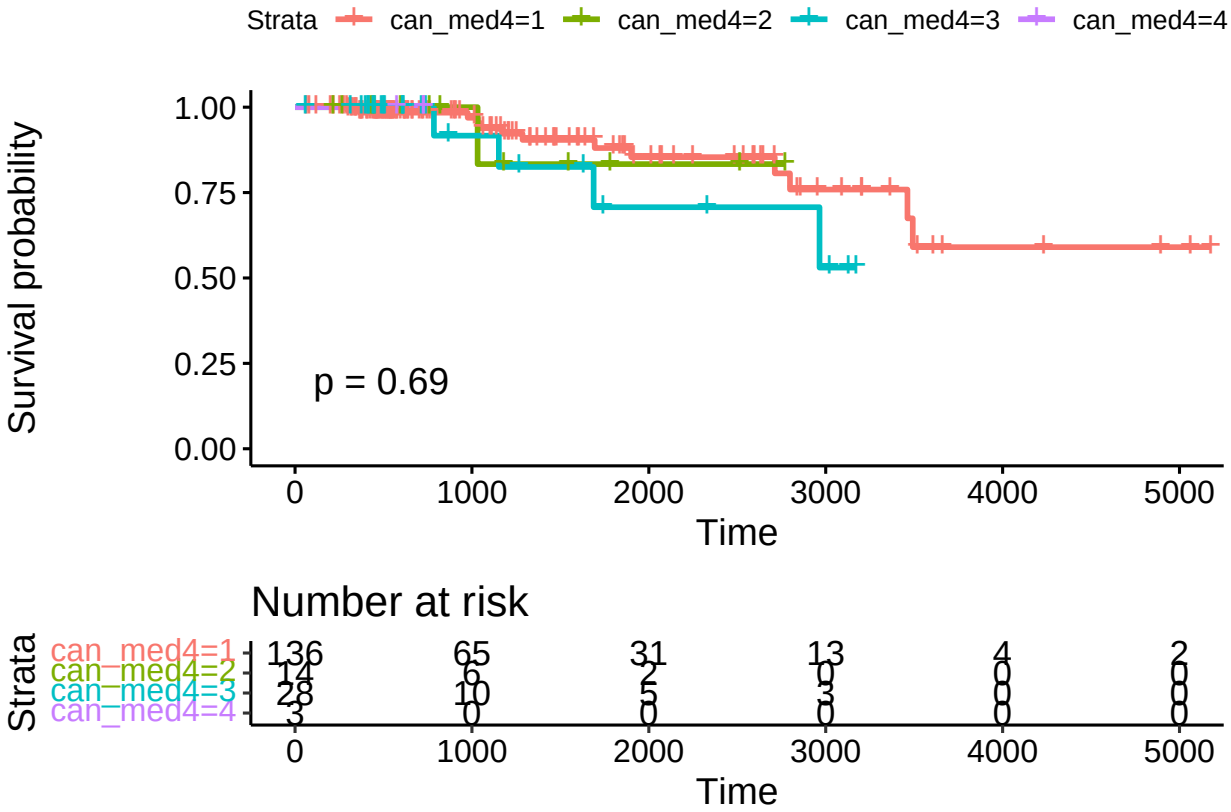
```
ggsurvplot(HC41,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS41,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

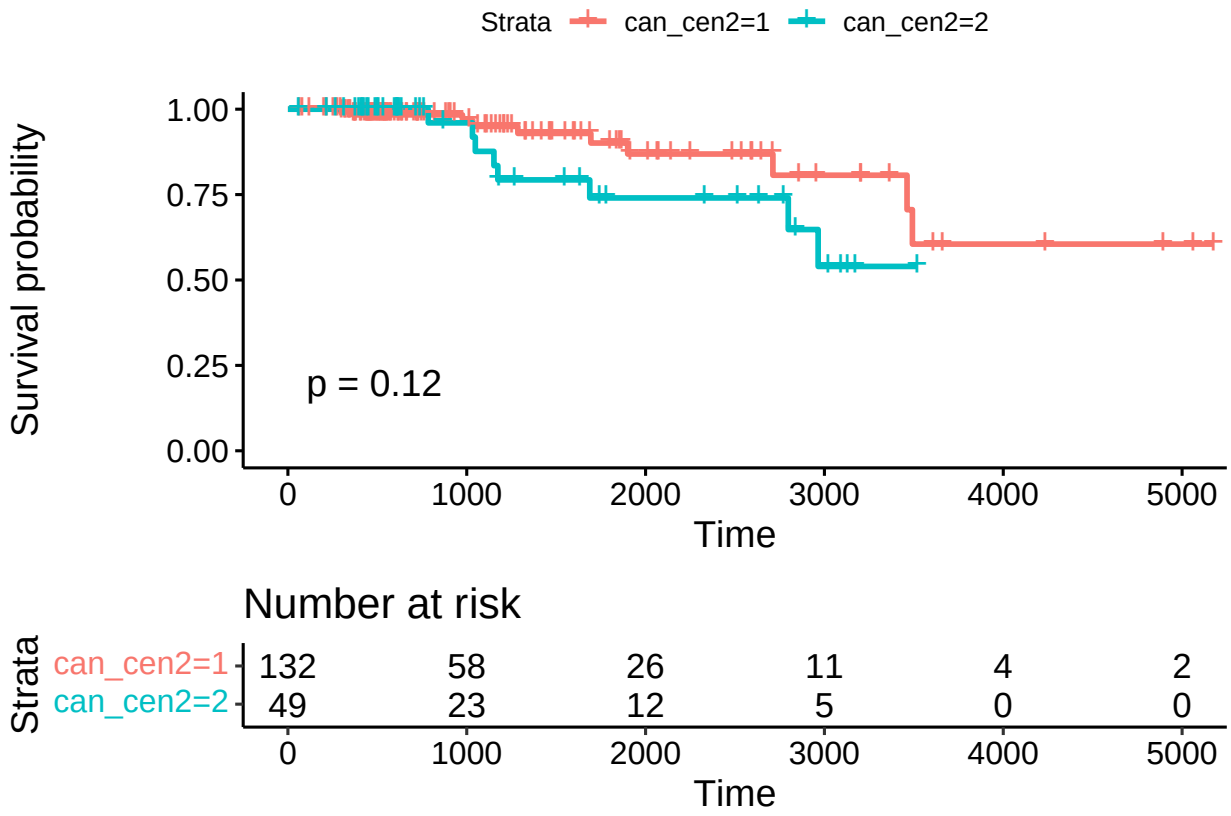



```
ggsurvplot(HW41,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

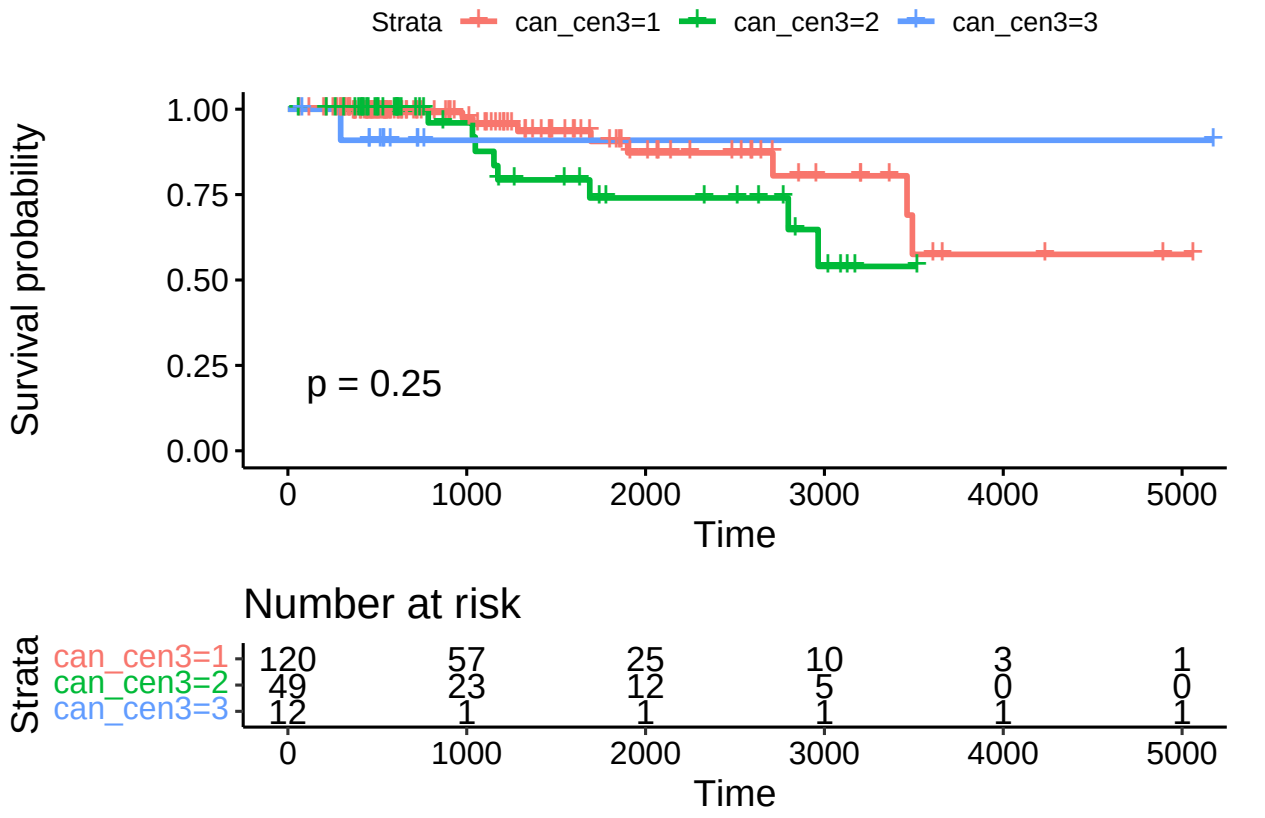


4.2.42. Canberra distance measurement + Centroid Linkage

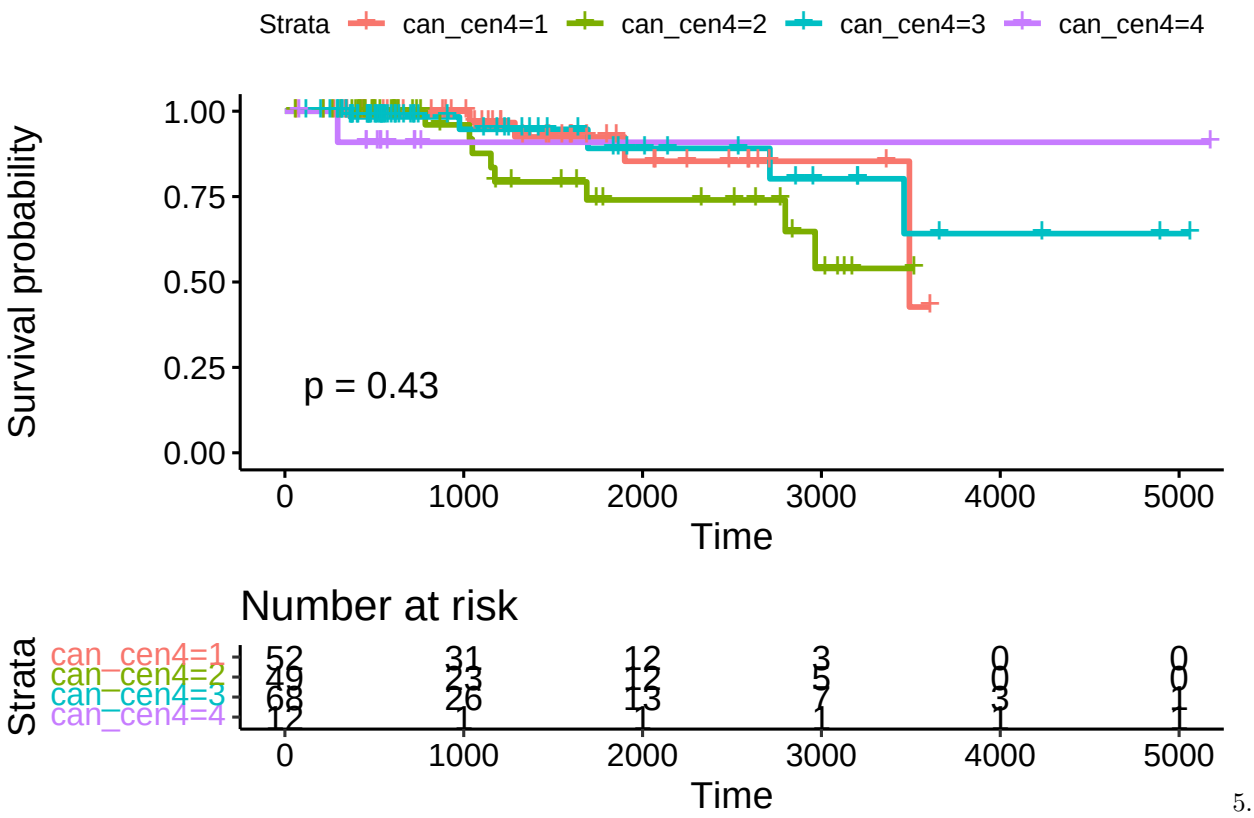
```
ggsurvplot(HC42,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HS42,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
ggsurvplot(HW42,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```



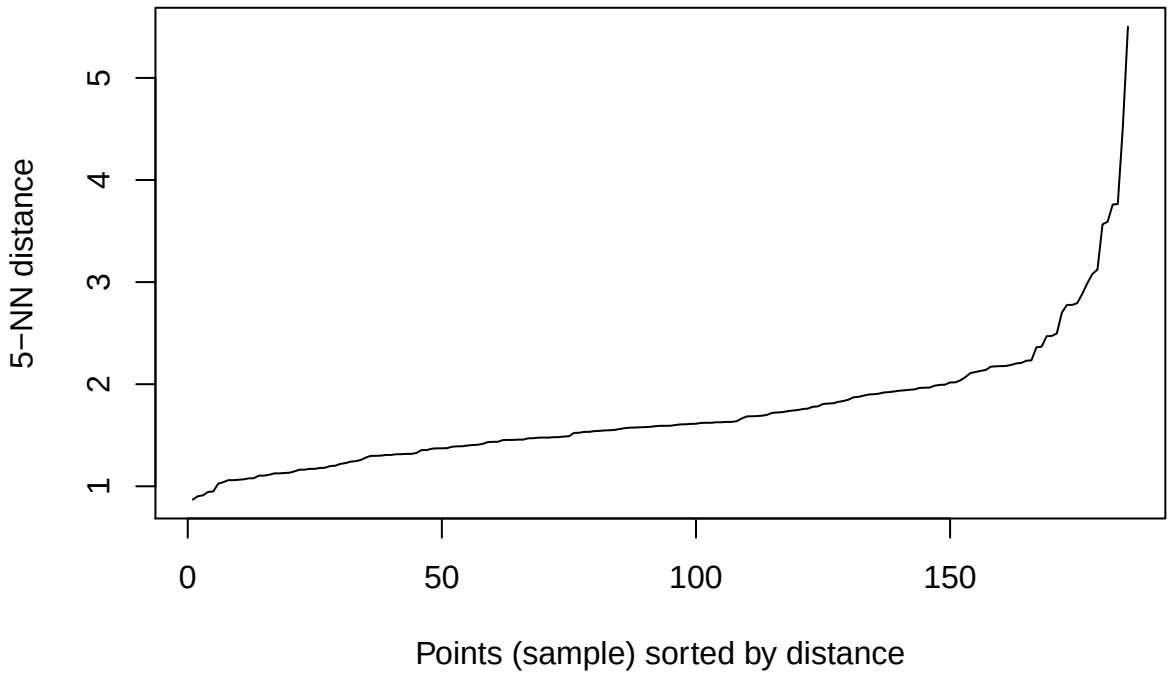
5.

DBSCAN 5.1. Obtain the best EPS

```
# DBSCAN
library("dbscan")
```

```
## Warning: package 'dbscan' was built under R version 3.6.2
```

```
kNNdistplot(tucker1,k=5)
abline(h=0.15,lty=2)
```



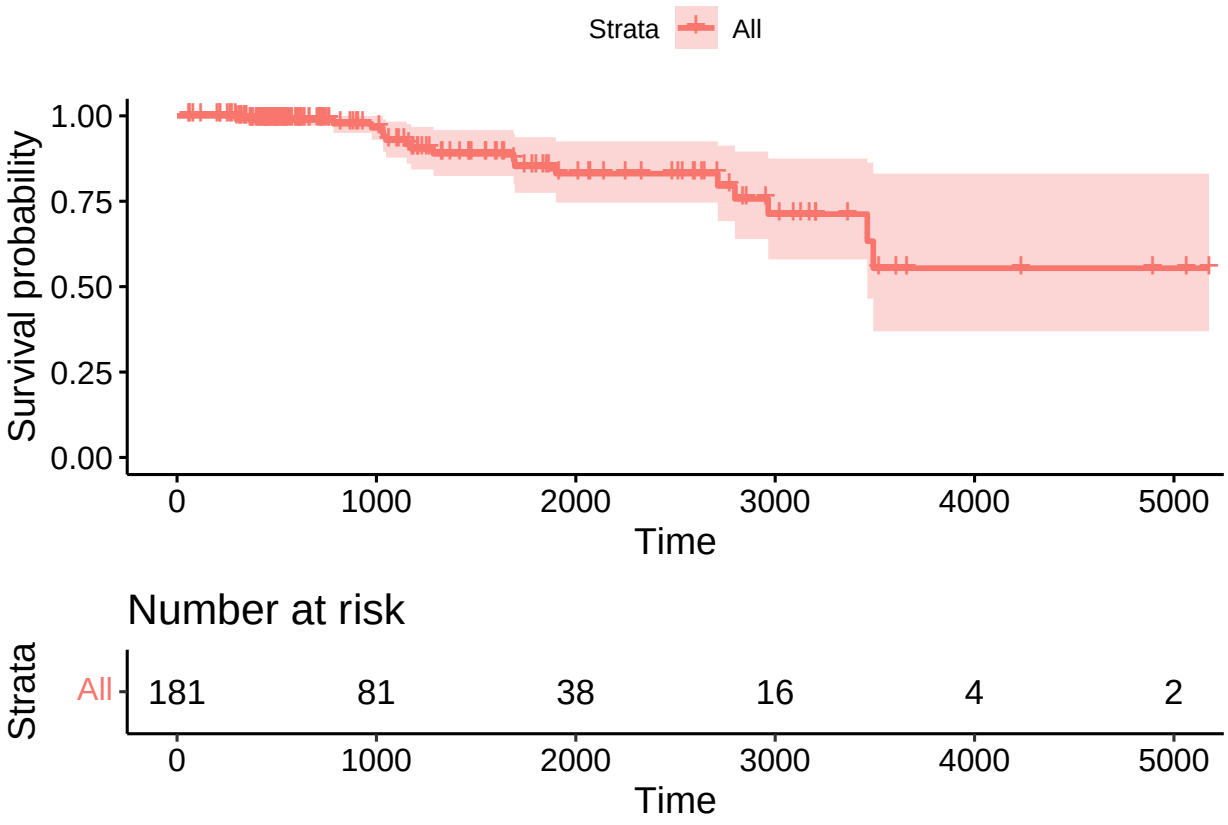
5.2. DBSCAN

```
db=dbscan(tucker1,eps=16,MinPts=20)
```

```
## Warning in dbscan(tucker1, eps = 16, MinPts = 20): converting argument MinPts
## (fpc) to minPts (dbscan)!
```

```
data1$db=db$cluster
dbl = survfit(obj ~ db, data = data1)
ggsurvplot(dbl,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

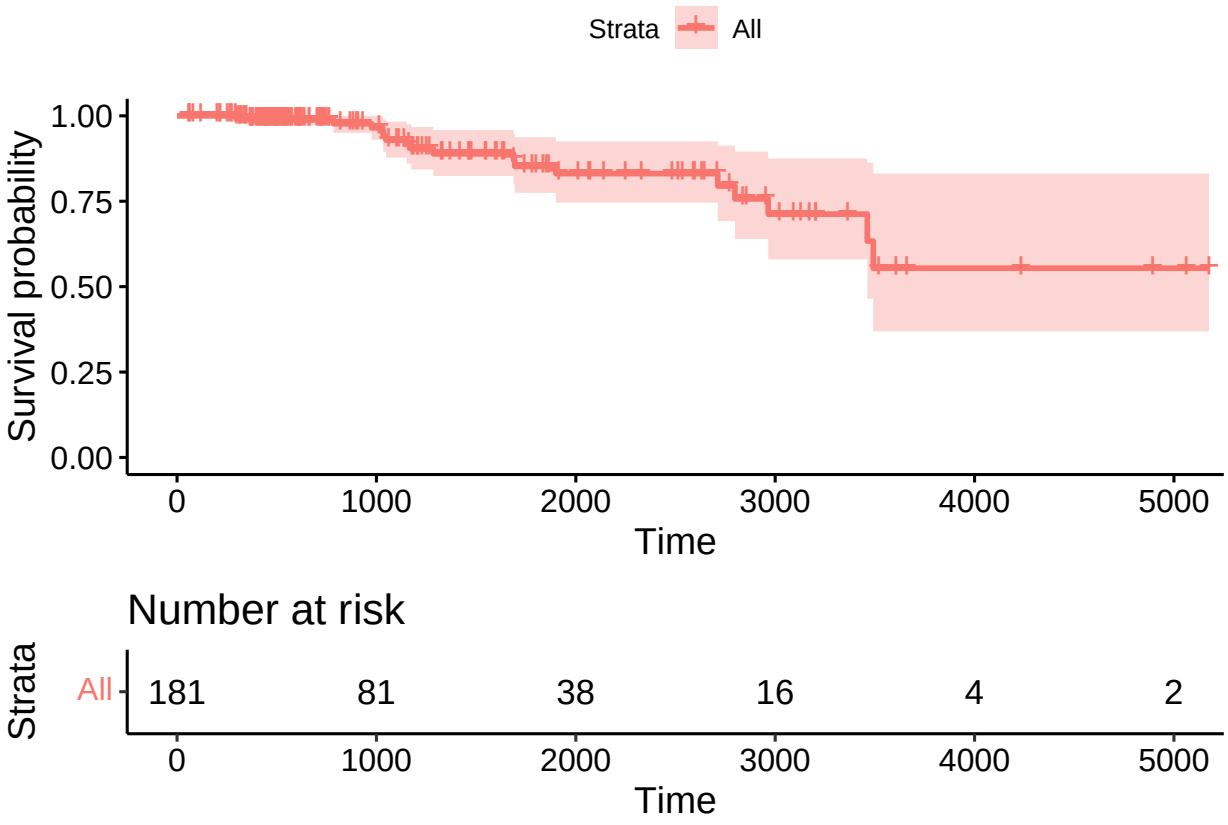
```
## Warning in .pvalue(fit, data = data, method = method, pval = pval, pval.coord = pval.coord, : There a
##  This is a null model.
```



5.3. hdbscan

```
hdb=hdbscan(tucker1,10)
data1$hdb=hdb$cluster
hdbl = survfit(obj ~ hdb, data = data1)
ggsurvplot(hdbl,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```

```
## Warning in .pvalue(fit, data = data, method = method, pval = pval, pval.coord = pval.coord, : There a
##  This is a null model.
```



6.EM algorithm

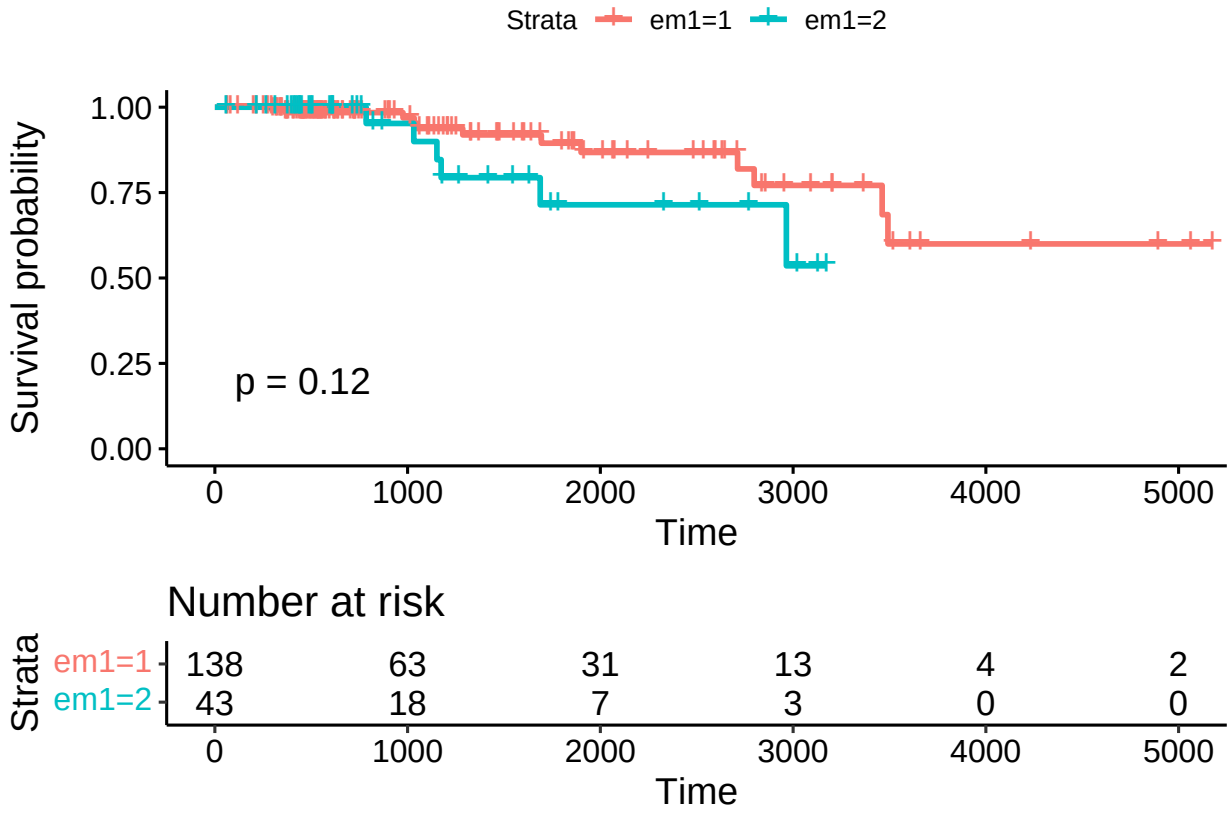
```
# GMM
library(mclust)
```

```
## Warning: package 'mclust' was built under R version 3.6.2
```

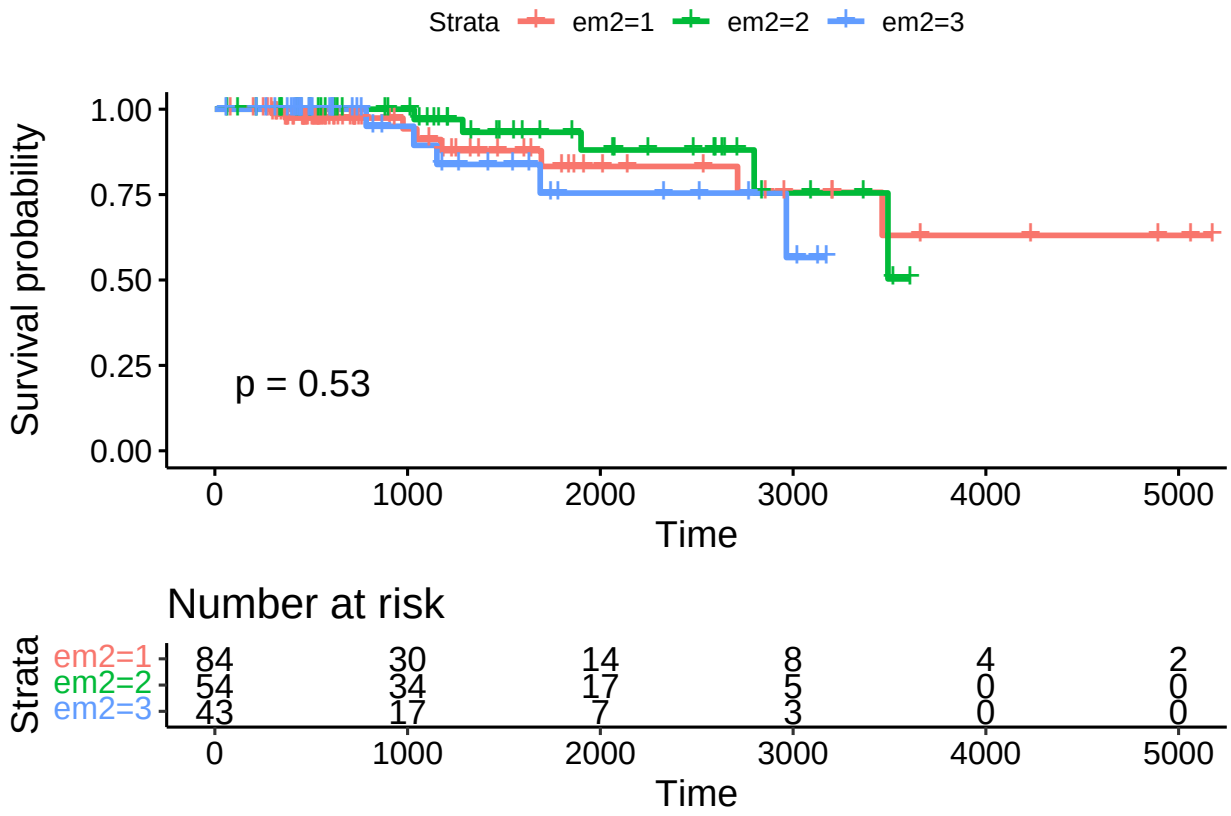
```
em1 = Mclust(tucker1,2)
em2 = Mclust(tucker1,3)
em3 = Mclust(tucker1,4)
data1$em1=em1$classification
data1$em2=em2$classification
data1$em3=em3$classification
```

6.1. Fit the results of EM algorithm

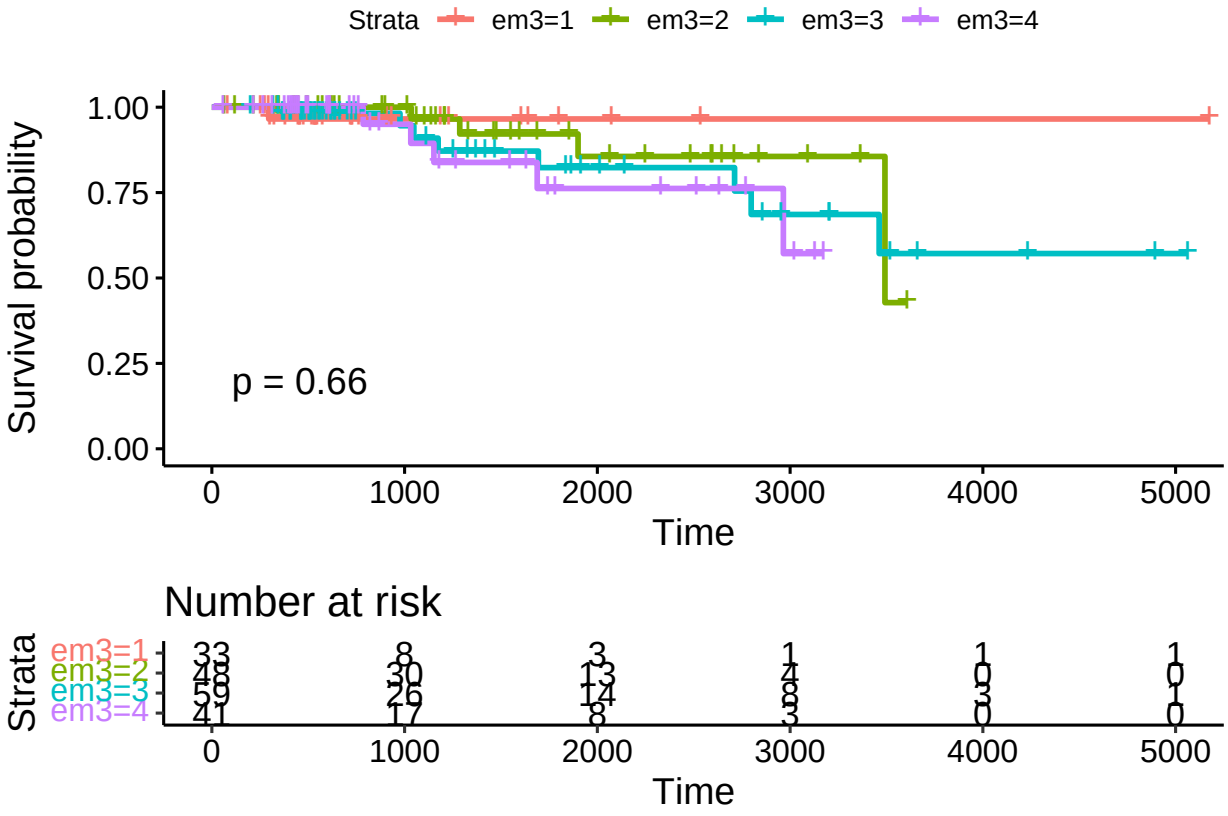
```
eml1 = survfit(obj ~ em1, data = data1)
ggsurvplot(eml1,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
eml2 = survfit(obj ~ em2, data = data1)
ggsurvplot(eml2,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```




```
eml3 = survfit(obj ~ em3, data = data1)
ggsurvplot(eml3,
           pval=T,
           risk.table=T,
           risk.table.height=.3)
```


Strata em3=1 em3=2 em3=3 em3=4
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.66
0
1000
2000
3000
4000
5000
Time
Number at risk
Strata
em3=1
em3=2
em3=3
em3=4
33 8 3 1 1 1
48 30 13 4 0 0
59 26 14 8 3 1
41 17 8 3 0 0
0
1000
2000
3000
4000
5000
Time