Exploratory Data Analysis

Bernabe Cano Paez

2025-11-24

Description of the dataset

The `actg` dataset can be found in the `GLDreg` package and comes from Hosmer et al. (2008). This data frame has 1151 rows and 16 columns.

```
data <- readRDS("actg_clean.rds")

dim(data)
```

```
## [1] 1151   16
```

```
#str(data)
#skimr::skim(data)
```

```
## id        : Patient identifier
## tx        : Treatment indicator (1 = IDV included, 0 = otherwise)
## txgrp     : Treatment group (zdv, idv_zdv, idv_d4t, d4t)
## sex       : Sex (male, female)
## ivdrug    : IV drug use history (never, currently, previously)
## hemophil  : Hemophilia indicator (1 = yes, 0 = no)
## karnof    : Karnofsky score functional status
##             (100 = Normal; no complaint no evidence of disease.
##               90 = Normal activity possible; minor signs/symptoms of disease.
##               80 = Normal activity with effort; some signs/symptoms of disease.
##               70 = Cares for self; normal activity/active work not possible.)
## priorzdv  : Months of prior ZDV use
## age       : Age at enrollment (years)
## cd4_lvl   : CD4 stratum at screening (0 <= 50, 1 >50)
## race      : Race/Ethnicity
##           (wnh = White Non-Hispanic.
##            bnh = Black Non-Hispanic.
##            h   = Hispanic.
##            api = Asian, Pacific Islander.
##            ai  = American Indian, Alaskan Native.).
## event     : Event indicator (1 = AIDS/death, 0 = censored)
## time_event: Days to AIDS diagnosis or death
## death     : Death indicator (1 = death, 0 = otherwise)
## time_death: Days to death
## base_cd4  : Baseline CD4 cell count (Cells/Milliliter)
```

Descriptive Statistics

The average and median event times are not very informative for most treatment groups. In particular, in the d4t and idv_d4t groups essentially no one has experienced the event, so the reported "mean" and "median" simply reflect small censored times and do not describe the event time.

Treatment Group	Number of Patients	Number of Events	Number of Censored Observations	Proportion Censored	Simple mean Time to Event	Median Time to Event
zdv	576	63	513	0.8906	223.8611	251.0
idv_zdv	572	33	539	0.9423	237.5262	263.0
d4t	1	0	1	1.0000	47.0000	47.0
idv_d4t	2	0	2	1.0000	42.5000	42.5

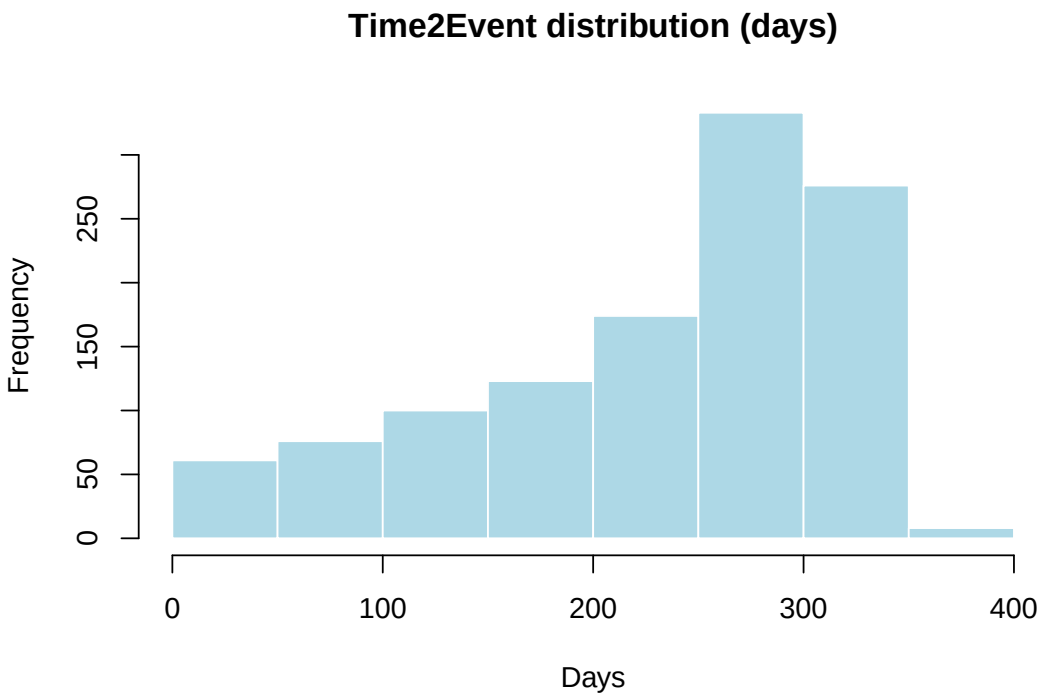
When aggregating treatment groups using only the binary variable tx (IDV included: yes/no), we observe slightly longer mean and median event times in the IDV group. However, this comparison is confounded by the fact that tx = 1 combines two regimens with very different censoring patterns (idv_zdv and idv_d4t). Because some of these regimens have almost no events, the aggregated descriptive statistics do not accurately reflect the underlying treatment effects. Therefore, the tx-based summary is less informative than the txgrp-based summary.

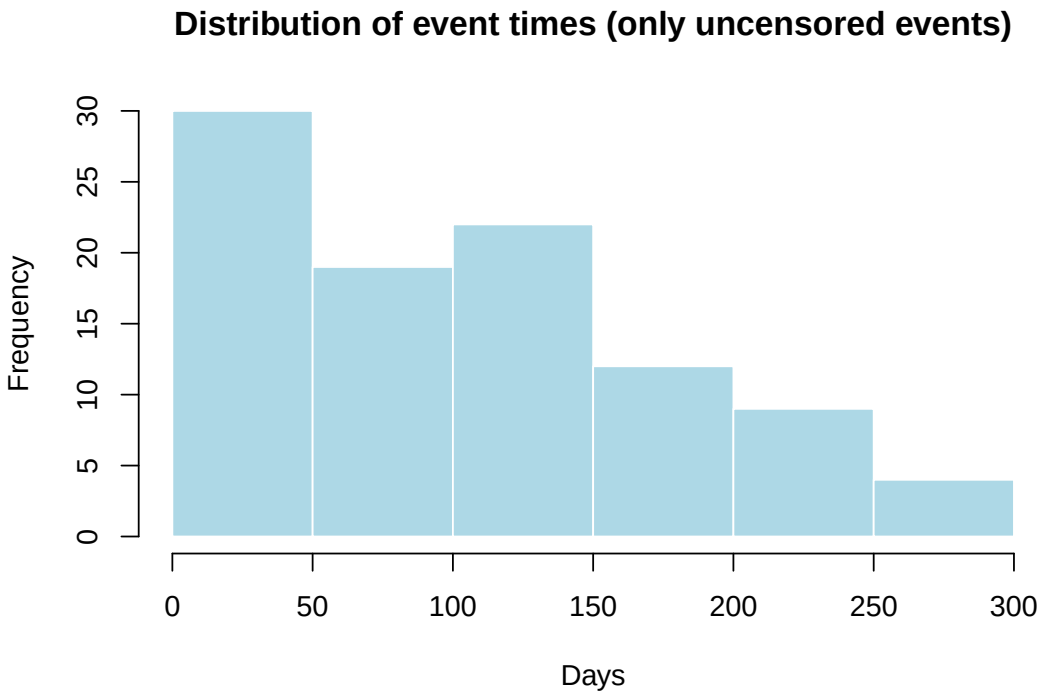
Consistent with Lecture 4, the mean and median event time are **not informative** for heavily censored groups.

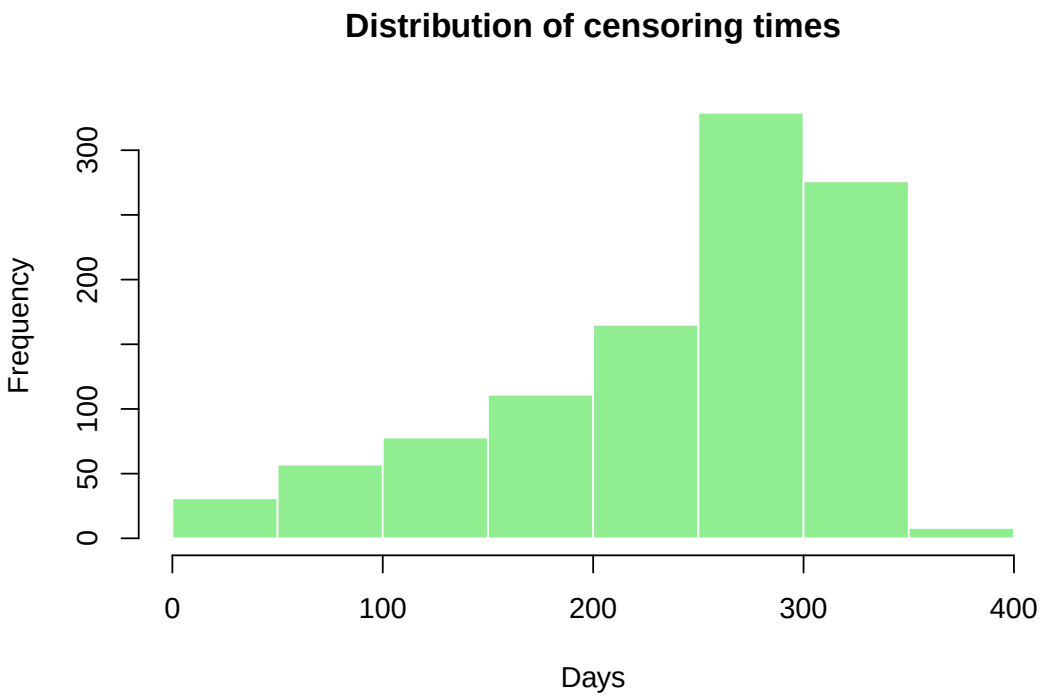
Treatment Group (Idv included: yes/no)	Number of Patients	Number of Events	Number of Censored Observations	Proportion Censored	Simple mean Time to Event	Median Time to Event
0	577	63	514	0.8908	223.5546	251
1	574	33	541	0.9425	236.8467	263

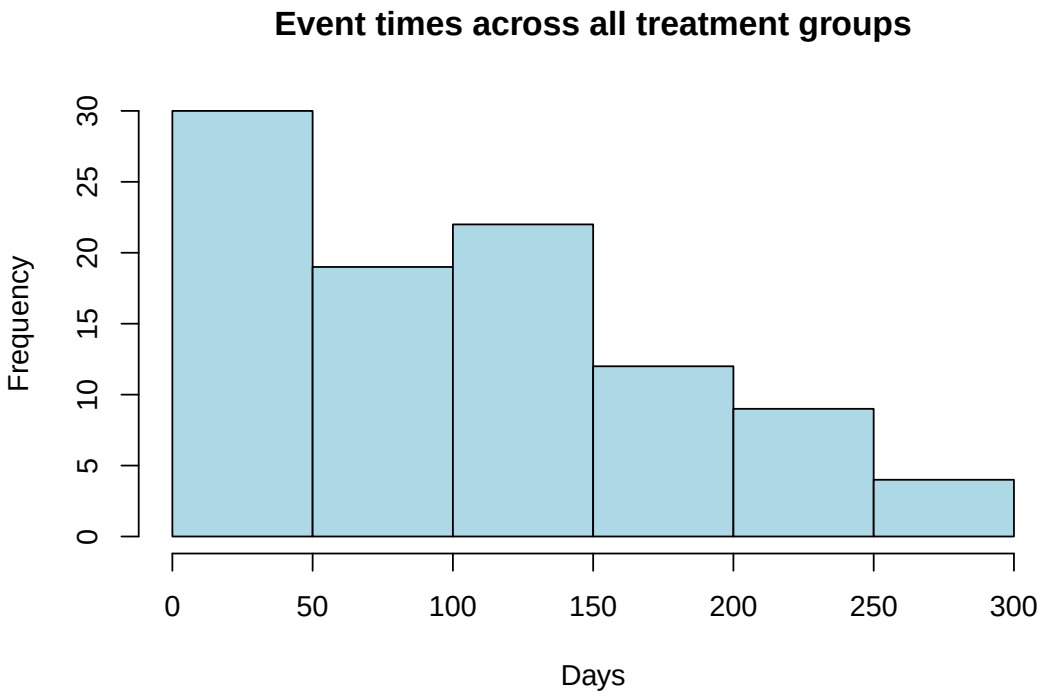
Basic counts for the categorical variables indicate that the sample is highly unbalanced across several characteristics. Most participants are male (951 vs. 200 females), the majority report no history of intravenous drug use (968 never vs. 179 previously vs. 4 currently), and roughly two-thirds of the patients fall into the higher CD4 stratum (>50). These imbalances suggest that covariate effects may play a role in the progression to AIDS or death and will need to be accounted for in later modeling.

Regarding treatment, the dataset provides two different codings: a four-level treatment group (txgrp) and a binary indicator for IDV inclusion (tx). The descriptive results highlight important differences between these two codings.

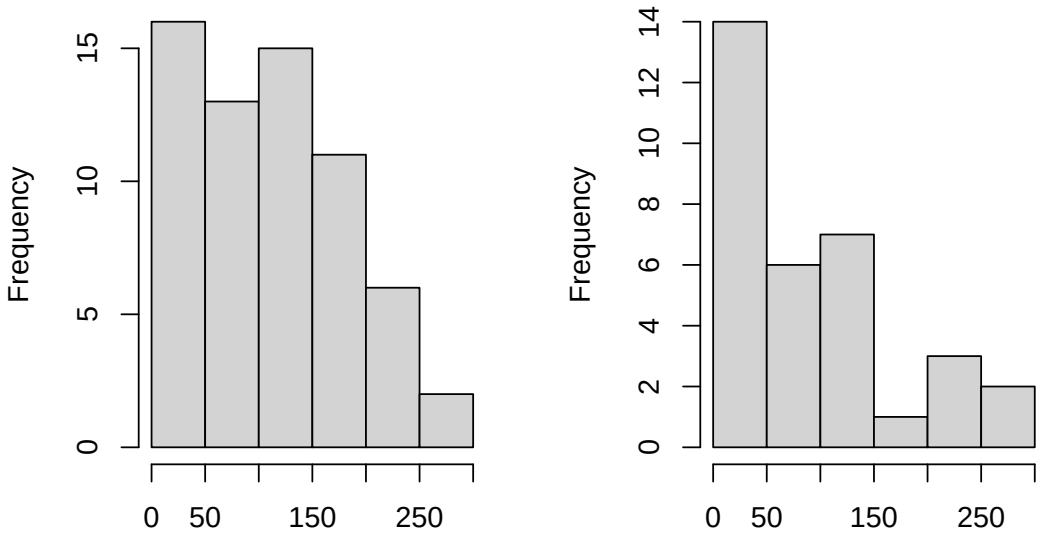

Time2Event distribution (days)
Frequency
0
50
150
250
0
100
200
300
400
Days


Distribution of event times (only uncensored events)
Frequency
0
5
10
15
20
25
30
0
50
100
150
200
250
300
Days

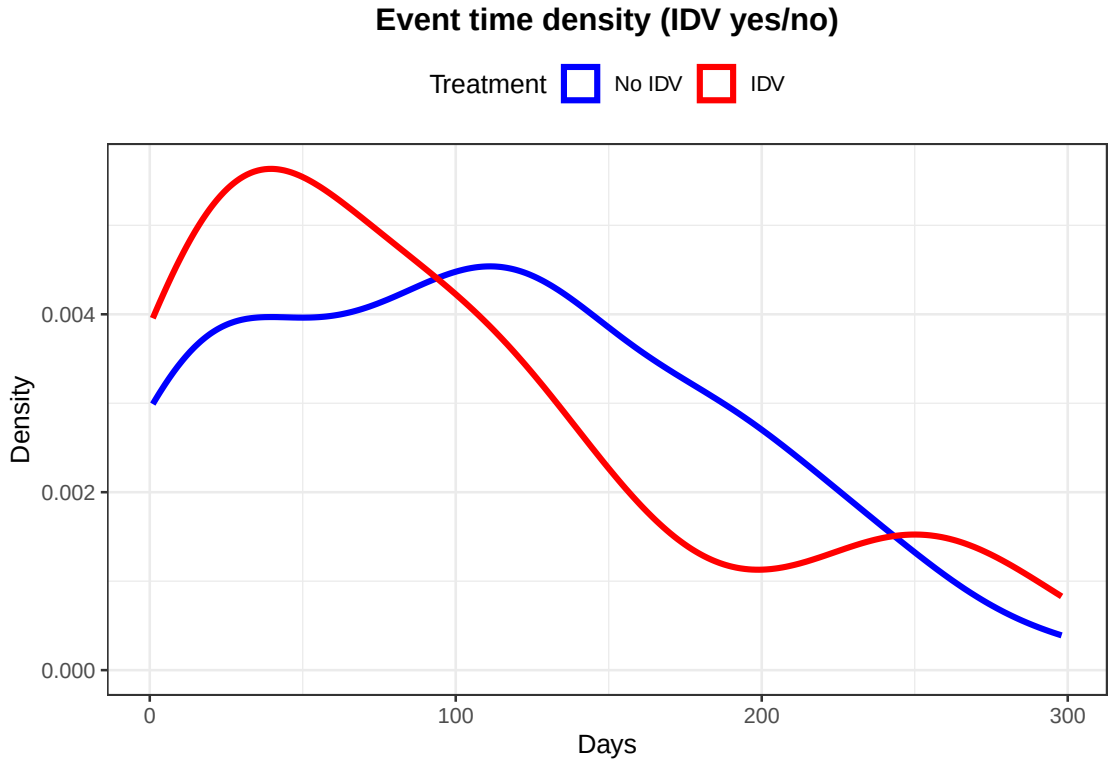

Distribution of censoring times
Frequency
0
50
100
200
300
0
100
200
300
400
Days


Event times across all treatment groups
Frequency
0
5
10
15
20
25
30
0
50
100
150
200
250
300
Days

ata$time_event[data$txgrp == "zdv"a$time_event[data$txgrp == "idv_zd

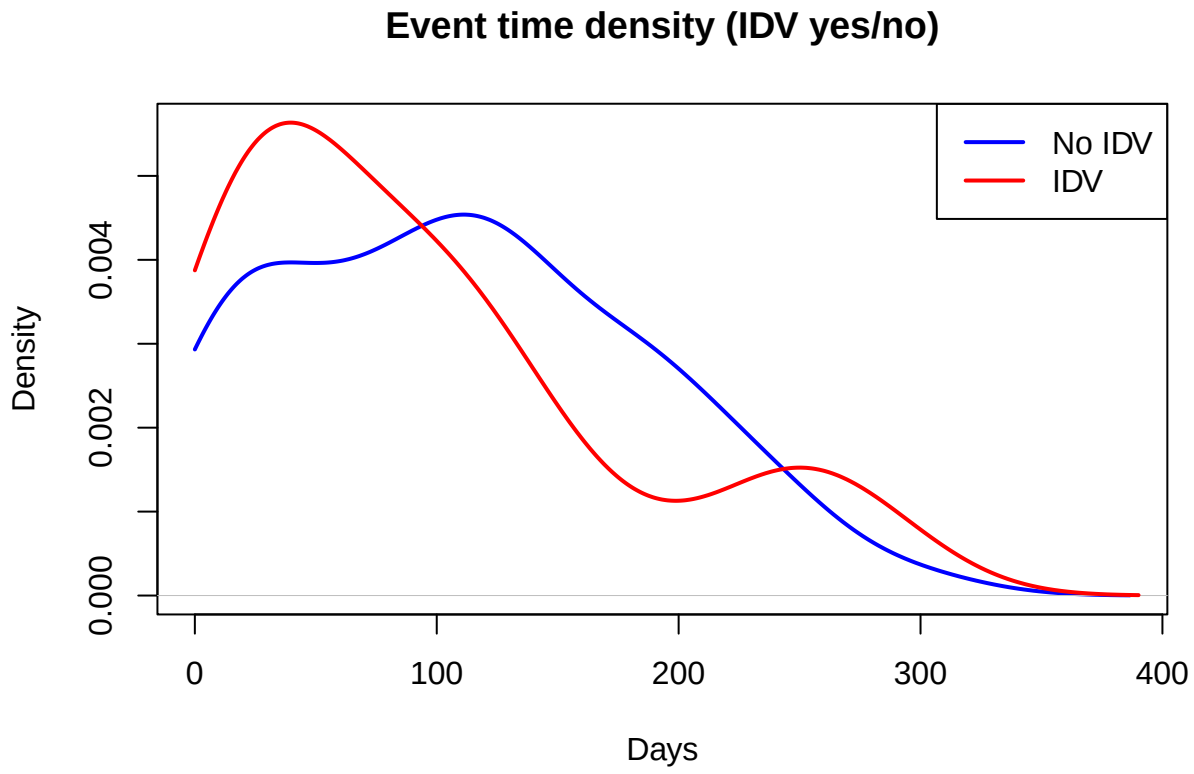


a$time_event[data$txgrp == "zdv" & data$evtime_event[data$txgrp == "idv_zdv" & data$

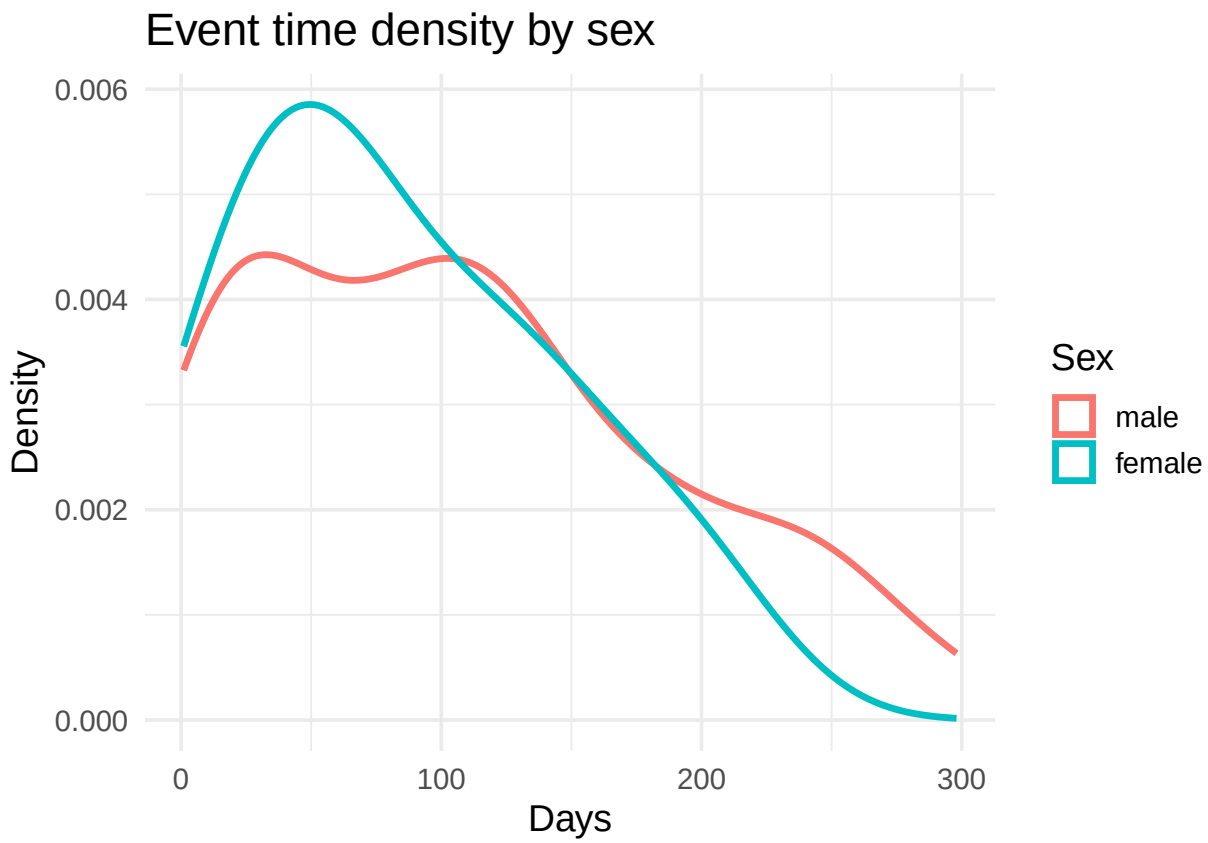


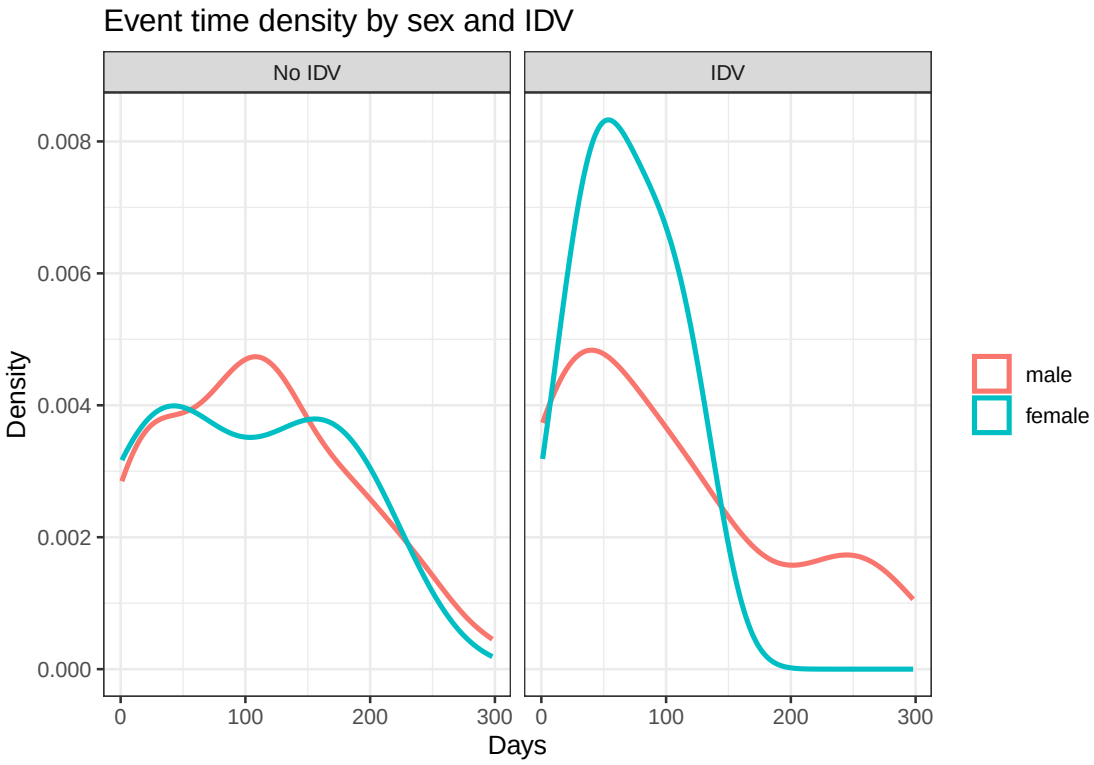

```
d0 <- density(data$time_event[data$tx == 0 & data$event == 1], from = 0)
d1 <- density(data$time_event[data$tx == 1 & data$event == 1], from = 0)
```

```r
plot(d0, col="blue", lwd=2,
     main="Event time density (IDV yes/no)",
     xlab="Days", ylim=c(0, max(d0$y, d1$y)))
lines(d1, col="red", lwd=2)
legend("topright", legend=c("No IDV","IDV"), col=c("blue","red"), lwd=2)
```




```r
ggplot(data %>% filter(event == 1),
       aes(x = time_event, color = sex)) +
  geom_density(size = 1.2) +
  labs(
    title = "Event time density by sex",
    x = "Days",
    y = "Density",
    color = "Sex"
  ) +
  theme_minimal(base_size = 14)
```


Event time density by sex
0.006
0.004
0.002
0.000
Density
0
100
200
300
Days
Sex
male
female

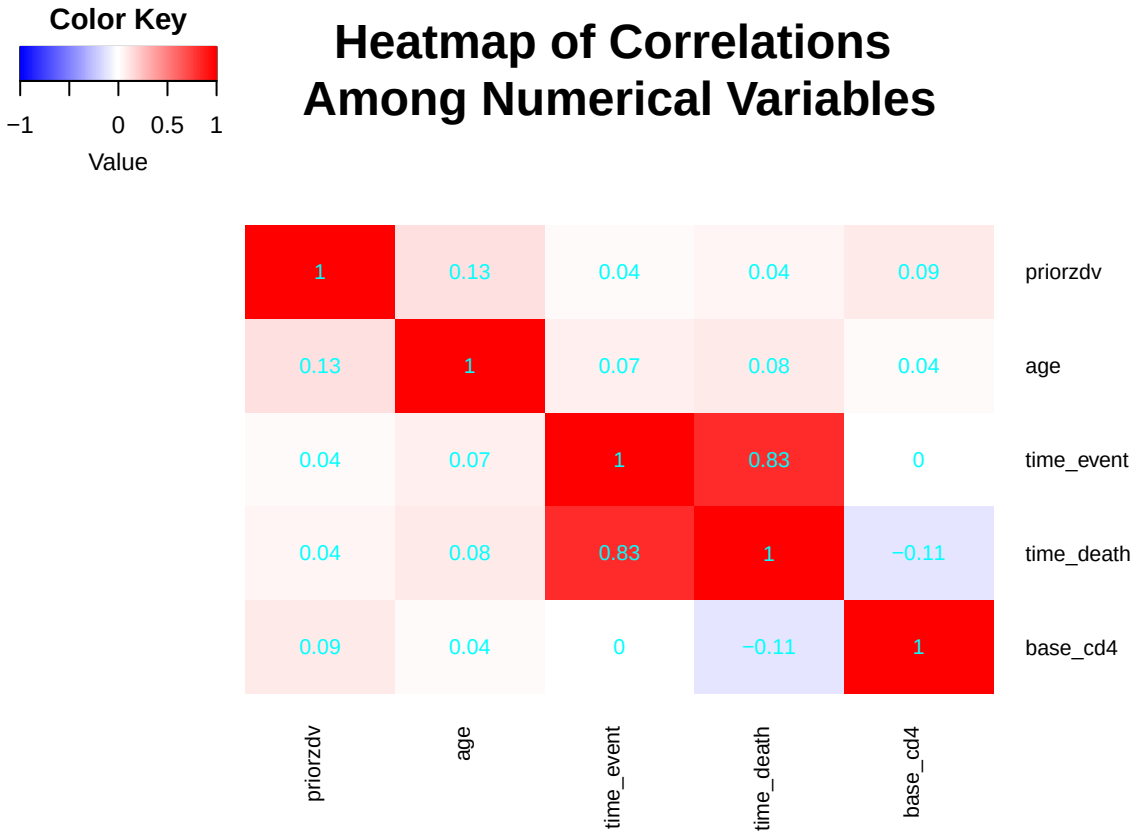

Event time density by sex and IDV
No IDV
IDV
0.008
0.006
0.004
0.002
0.000
Density
0
100
200
300
Days
male
female

Correlations

- **time_event** and **time_death** are positively correlated, reflecting the structural relationship: death is one type of event and shares follow-up time structure.

- **priorzdv** shows weak correlation with **base_cd4** and age, suggesting that ZDV exposure is not systematically linked to these baseline measures.

- Correlations among numeric variables are relatively modest, with no evidence of perfect linear associations.
- Baseline CD4 levels exhibit moderate correlation with time-to-death variable (≈ -0.11) and no correlation with time-to-event(≈ 0.004), consistent with clinical expectations.



Because the three outcome groups represent fundamentally different clinical trajectories, the correlation between baseline CD4 and time-to-event measures must be evaluated separately within each subgroup. A single overall correlation is not interpretable, as it conflates distinct event mechanisms, censoring structures, and risk pathways.

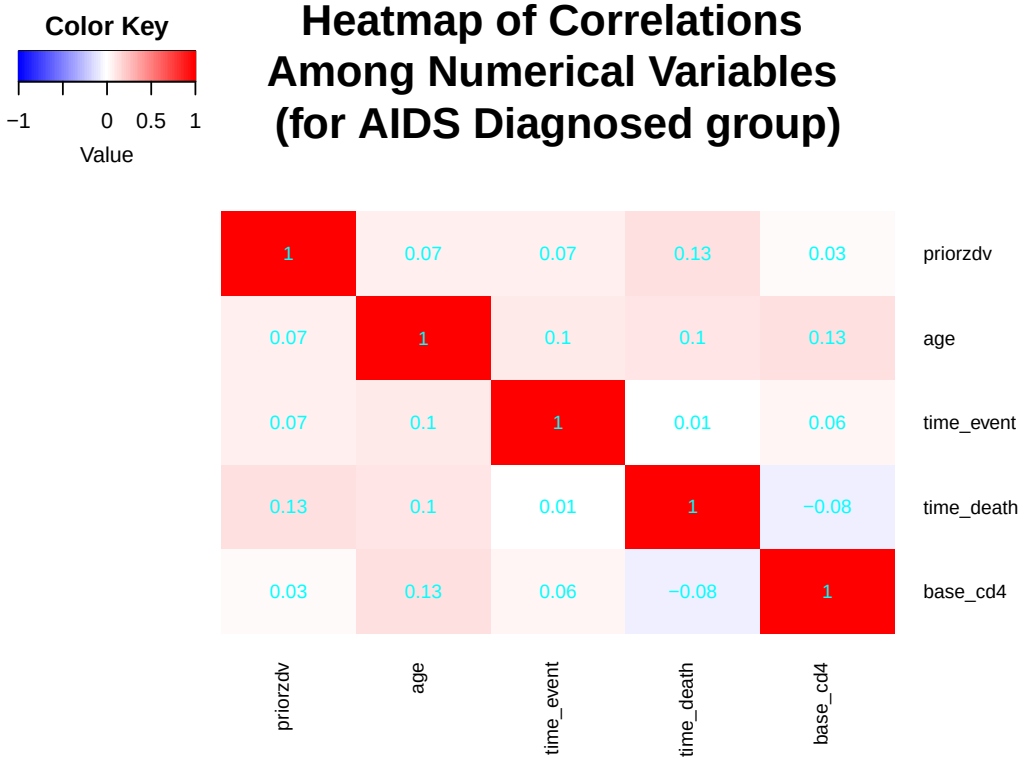
1. **AIDS Diagnosed subgroup** (event = 1, death = 0).
   For individuals who developed AIDS but did not die during follow-up, the relevant endpoint is the time to AIDS diagnosis. In this subgroup, `time_event` represents the actual observed progression time. It is clinically meaningful to assess whether lower baseline CD4 counts are associated with earlier progression to AIDS, which is highly plausible given the role of CD4 as a marker of immunological decline. Correlating baseline CD4 with time_event therefore directly examines whether impaired immune function predicts more rapid clinical deterioration.

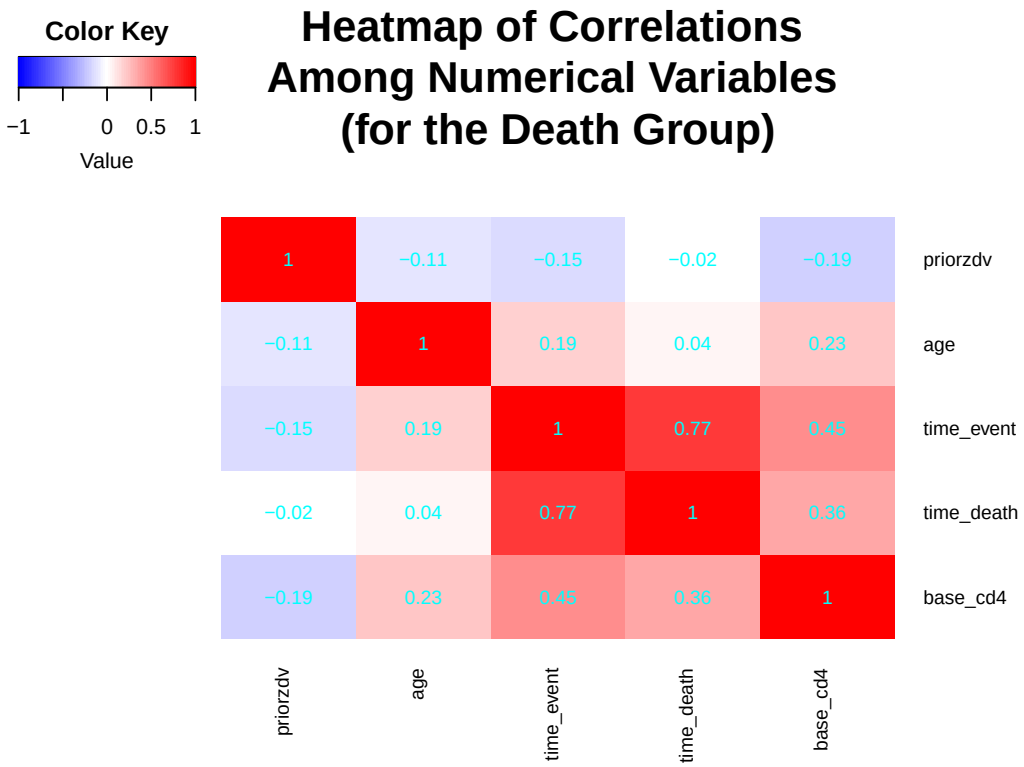
2. **Death subgroup** (death = 1).
   Among participants who died, the appropriate endpoint is time to death (time_death). This variable reflects the true failure time for this subgroup and captures the progression toward mortality. Exploring the correlation between baseline CD4 and time_death is clinically justified because lower CD4 levels are known to increase vulnerability to AIDS-related and opportunistic complications that may lead to death. A negative association (lower CD4 → shorter survival) would be consistent with established clinical understanding.

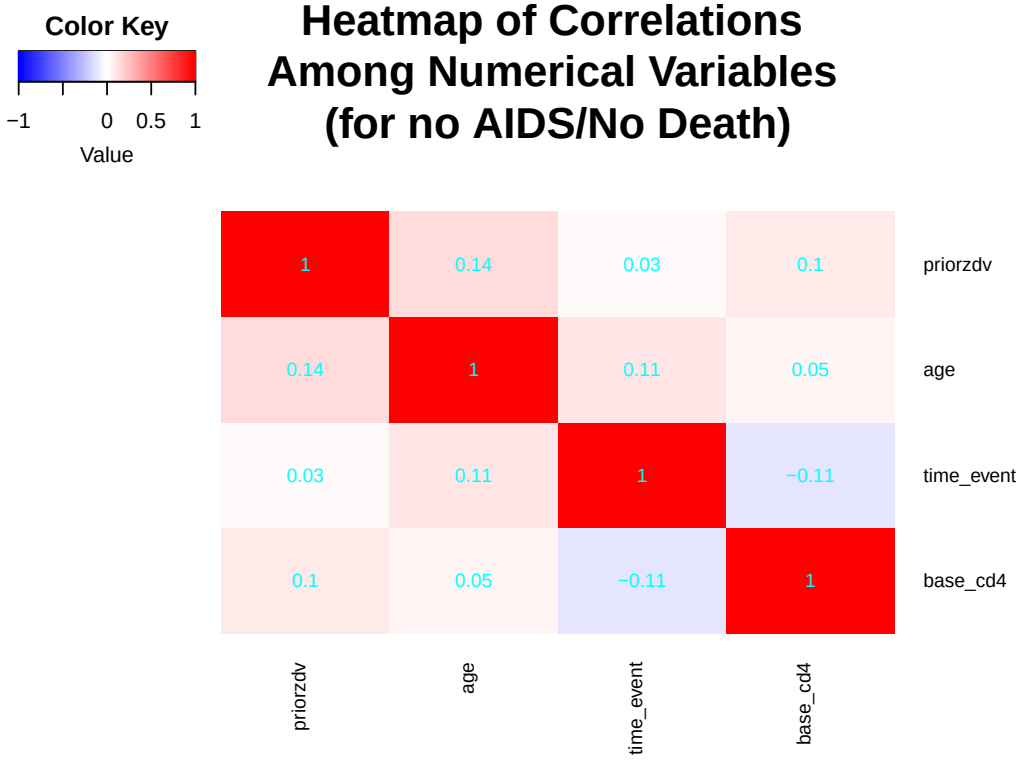
3. **NoDeath/NoDiagnosed subgroup** (event = 0, death = 0).
   For individuals who neither died nor developed AIDS, both time_event and time_death are purely censored and therefore do not represent meaningful biological times. The only interpretable time measure in this subgroup is the overall length of follow-up (time). Examining its correlation with baseline CD4 allows us to determine whether immune status is associated with the duration of observation prior to censoring. While a strong association is not expected, this exploration confirms whether follow-up time varies systematically with baseline immune function. Overall, analyzing time–CD4 correlations within outcome-specific strata ensures that each correlation reflects a coherent clinical process associated with a real, observed event (AIDS diagnosis or death) or with meaningful follow-up duration for censored individuals. This stratified approach avoids mixing incompatible event types and preserves the interpretability of both the clinical and statistical relationships.

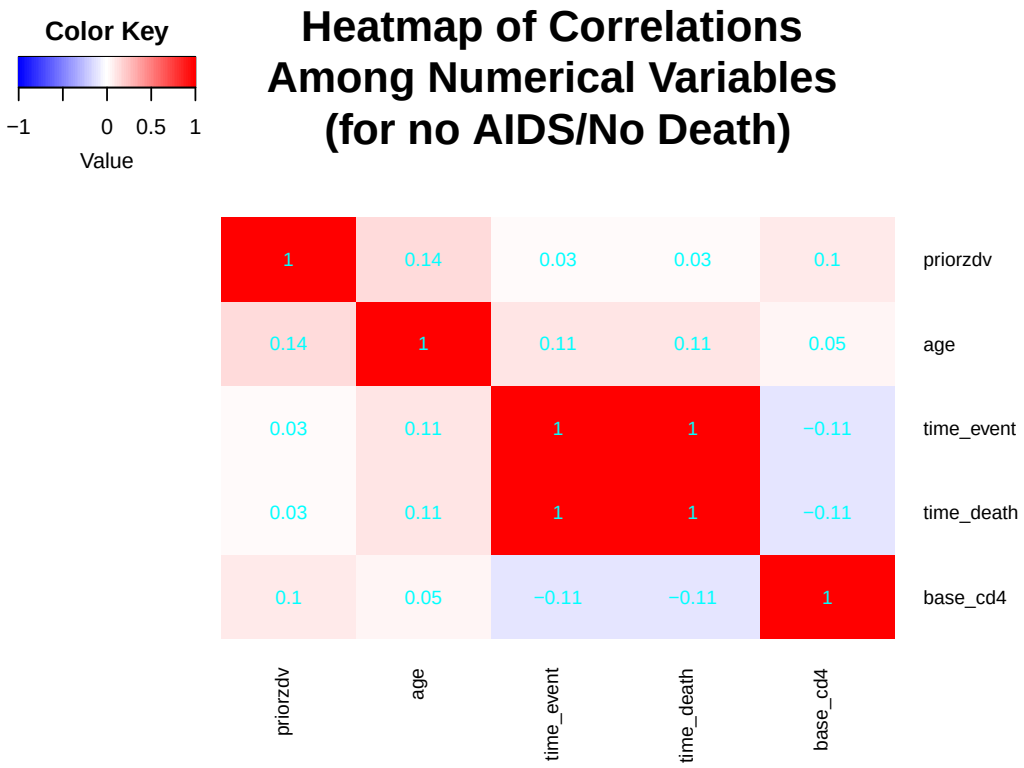
Overall, analyzing time–CD4 correlations within outcome-specific strata ensures that each correlation reflects a coherent clinical process associated with a real, observed event (AIDS diagnosis or death) or with meaningful follow-up duration for censored individuals. This stratified approach avoids mixing incompatible event types and preserves the interpretability of both the clinical and statistical relationships.

```
##
##      AIDS/Diagnosed               Death NoDeath/NoDiagnosed
##                  70                  26                1055
```



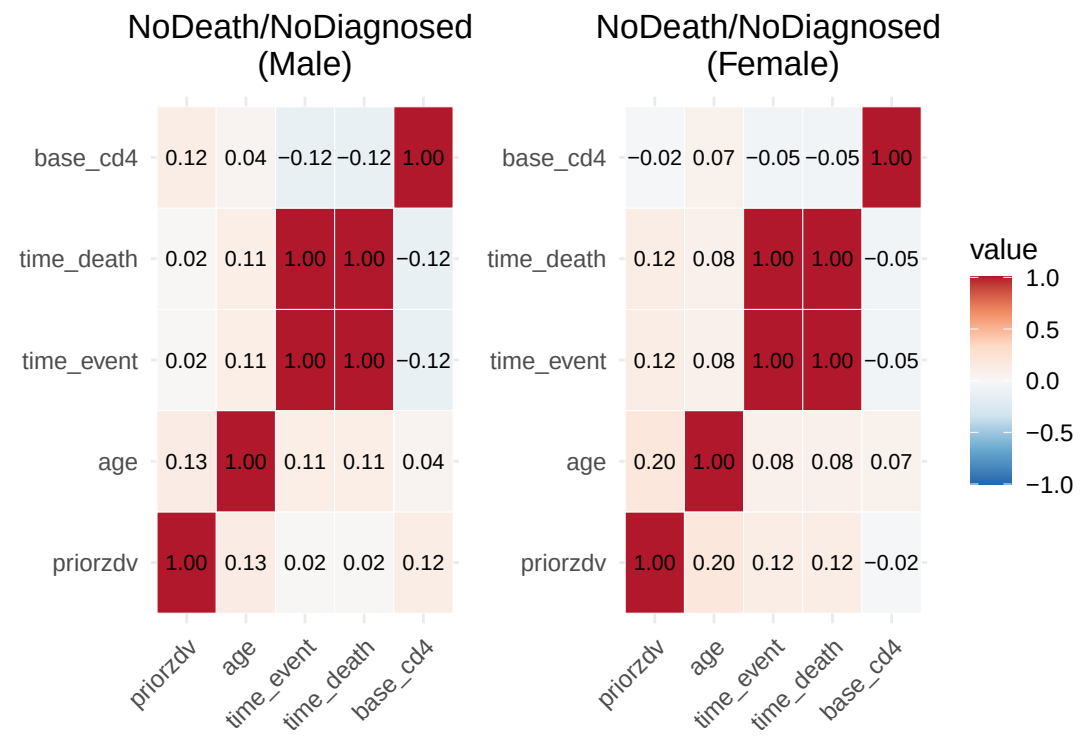

Color Key
−1
0
0.5
1
Value
Heatmap of Correlations Among Numerical Variables (for the Death Group)
1
−0.11
−0.15
−0.02
−0.19
priorzdv
−0.11
1
0.19
0.04
0.23
age
−0.15
0.19
1
0.77
0.45
time_event
−0.02
0.04
0.77
1
0.36
time_death
−0.19
0.23
0.45
0.36
1
base_cd4
priorzdv
age
time_event
time_death
base_cd4

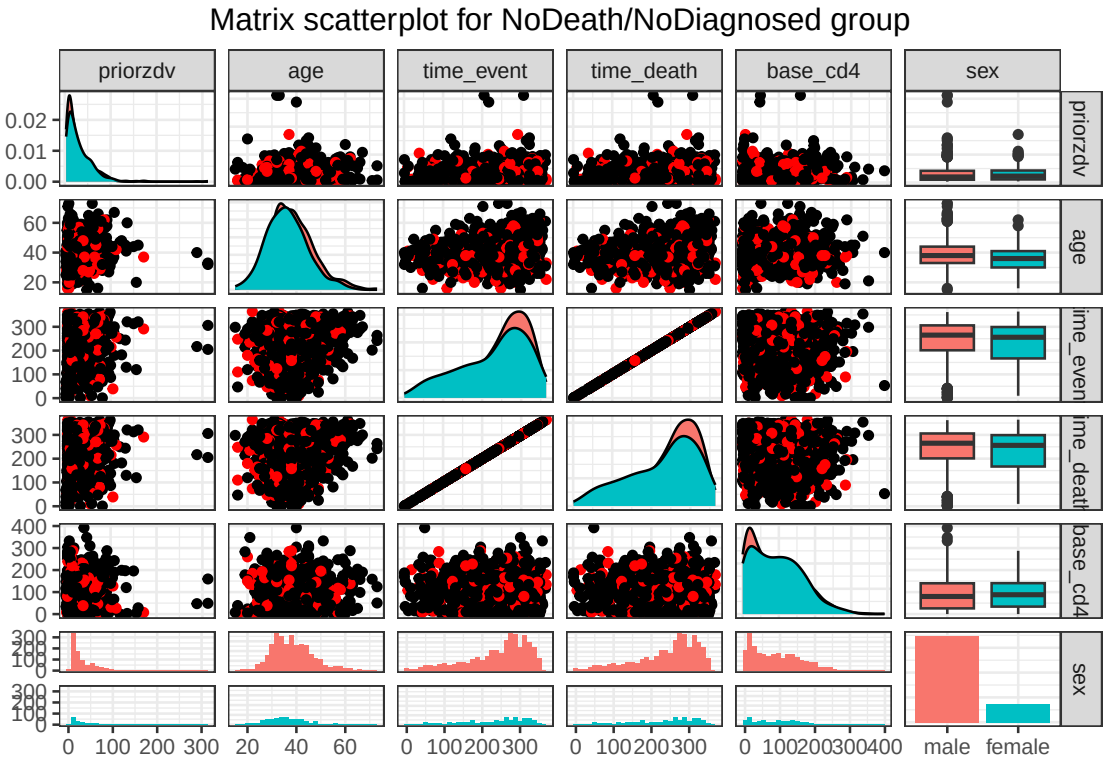

Color Key
−1
0
0.5
1
Value
Heatmap of Correlations Among Numerical Variables (for no AIDS/No Death)
1
0.14
0.03
0.1
priorzdv
0.14
1
0.11
0.05
age
0.03
0.11
1
−0.11
time_event
0.1
0.05
−0.11
1
base_cd4
priorzdv
age
time_event
base_cd4



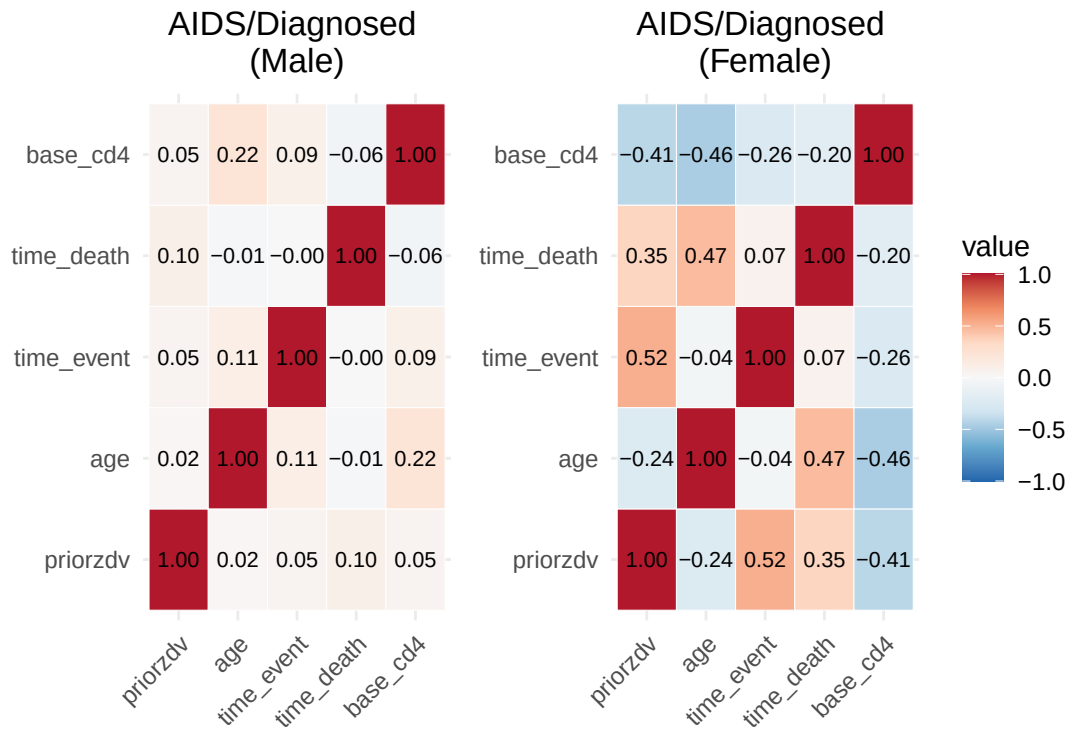
Correlations based on Sex and outcome (AIDS/Death/NoAIDSNoDeath)

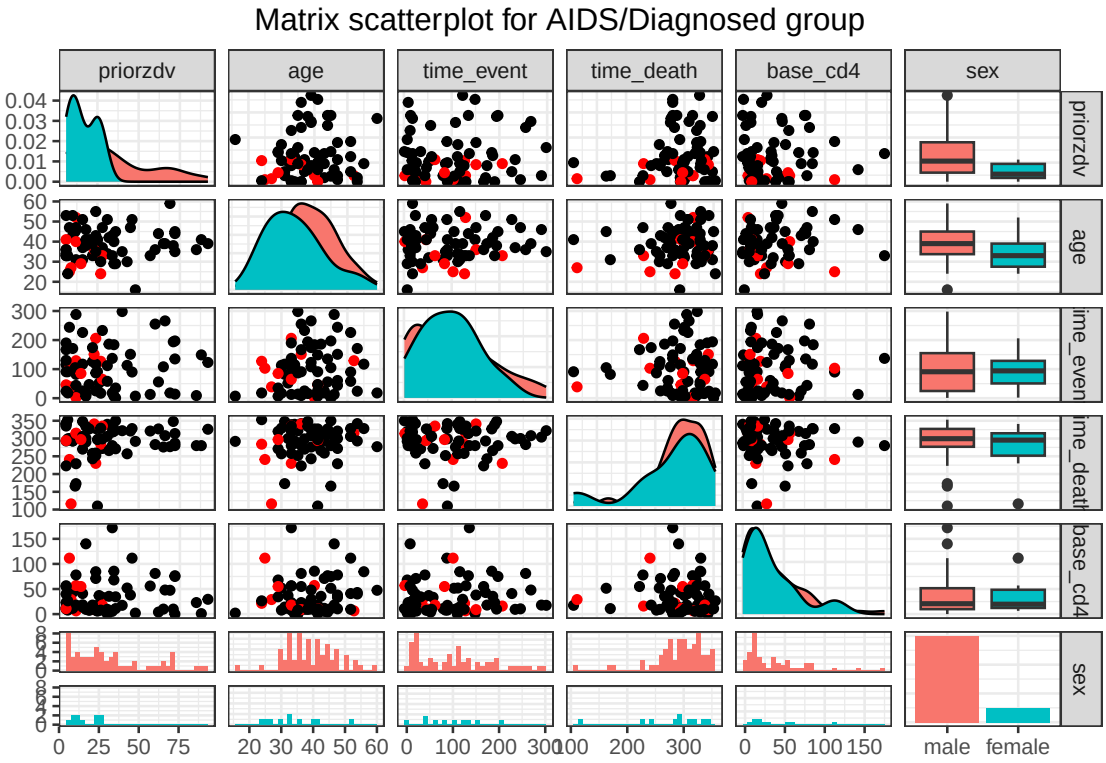
```
## [1] "NoDeath/NoDiagnosed"
```



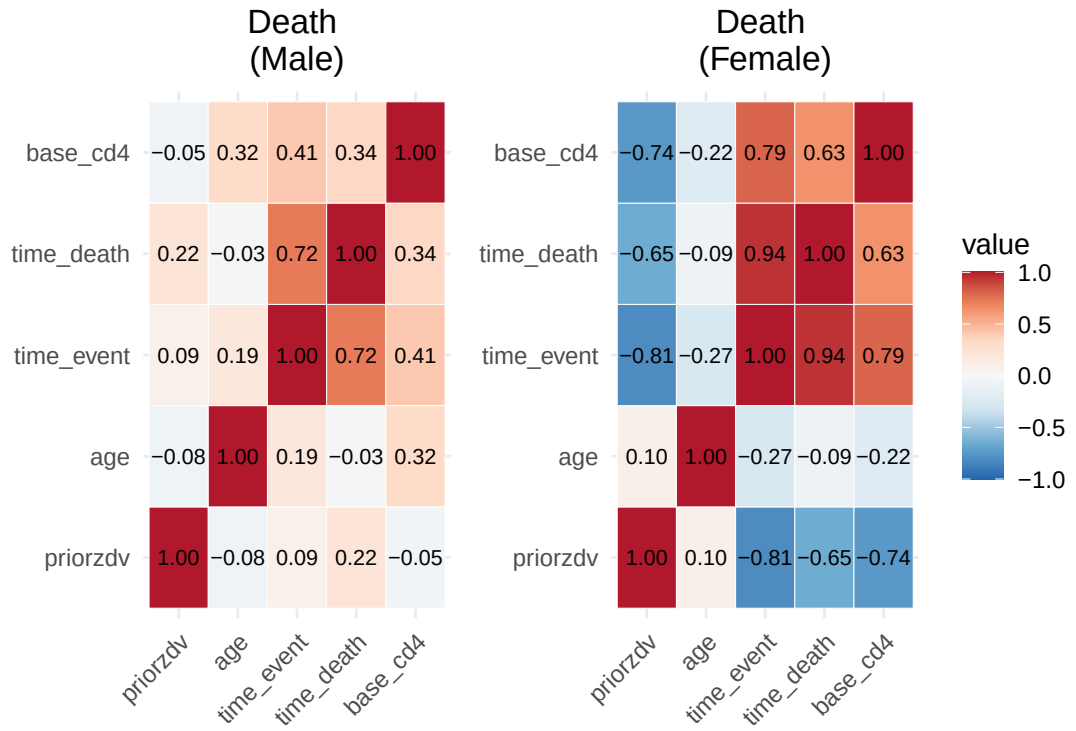


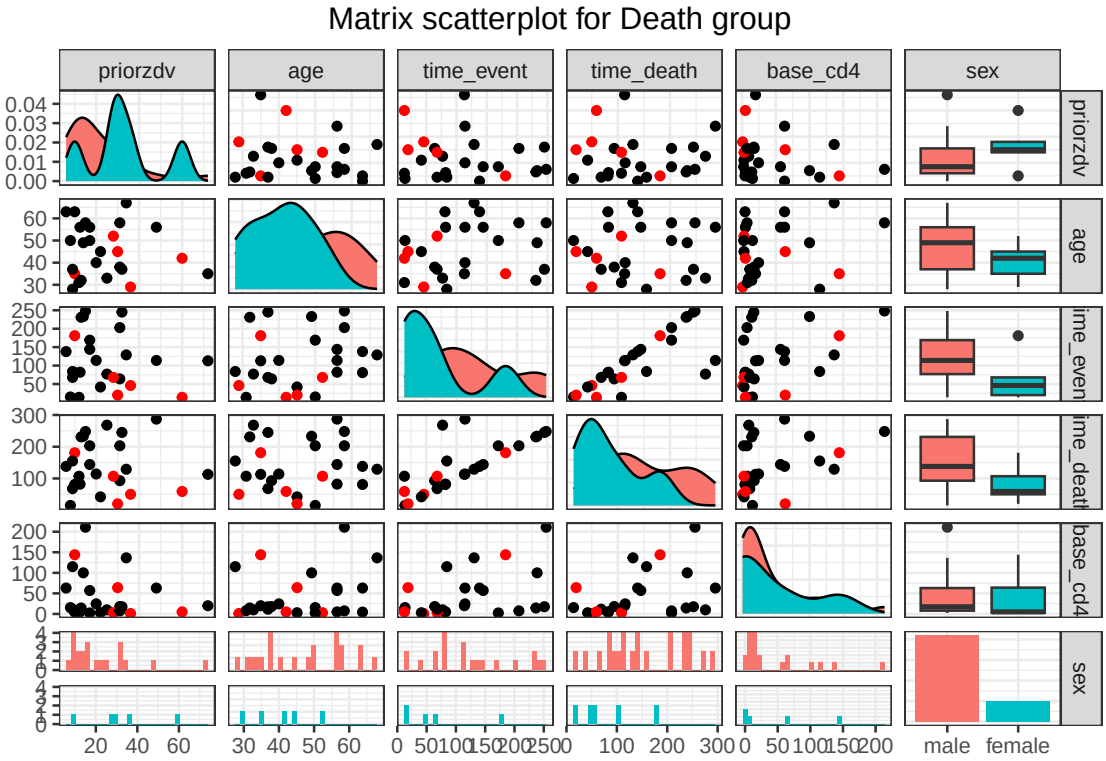

```
## [1] "AIDS/Diagnosed"
```


AIDS/Diagnosed (Male)
AIDS/Diagnosed (Female)
base_cd4
time_death
time_event
age
priorzdv
value
1.0
0.5
0.0
−0.5
−1.0


Matrix scatterplot for AIDS/Diagnosed group
priorzdv
age
time_event
time_death
base_cd4
sex
male
female

```
## [1] "Death"
```

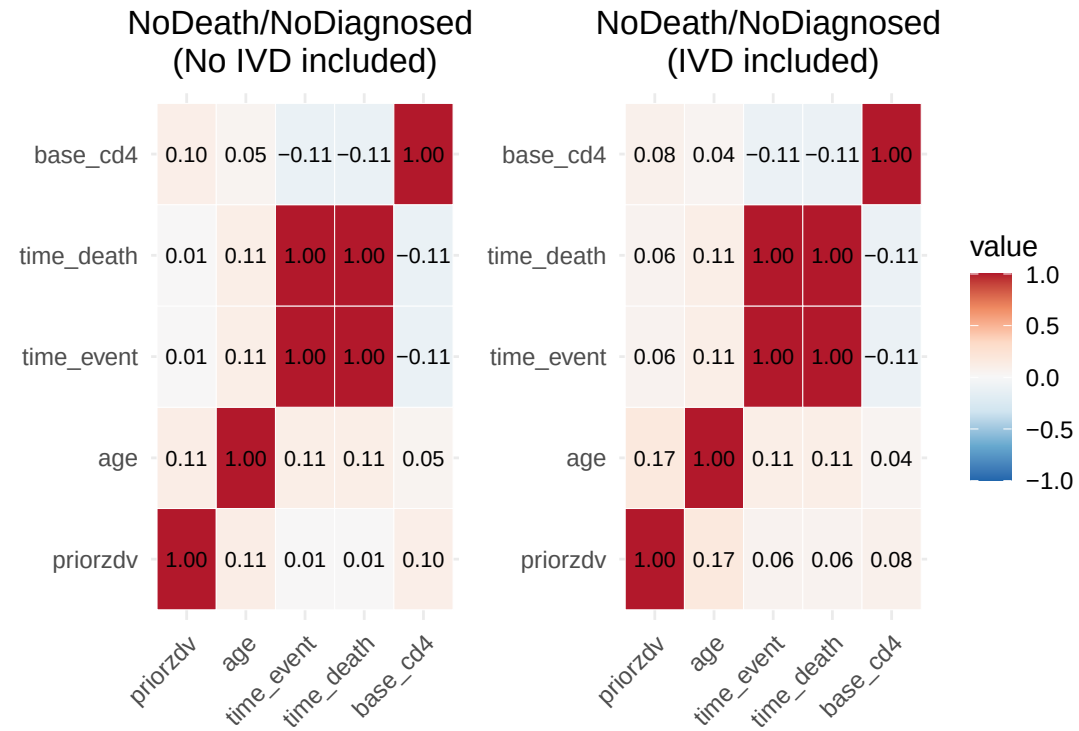


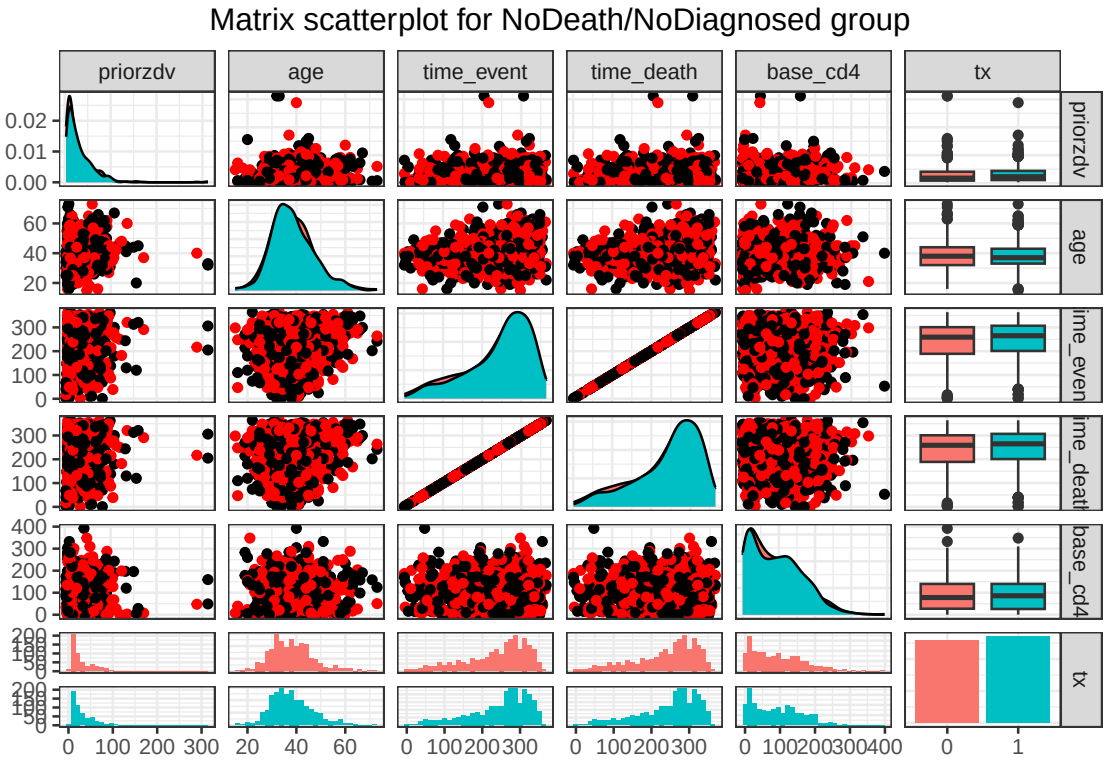


	AIDS/Diagnosed	Death	NoDeath/NoDiagnosed
male	60	21	870
female	10	5	185

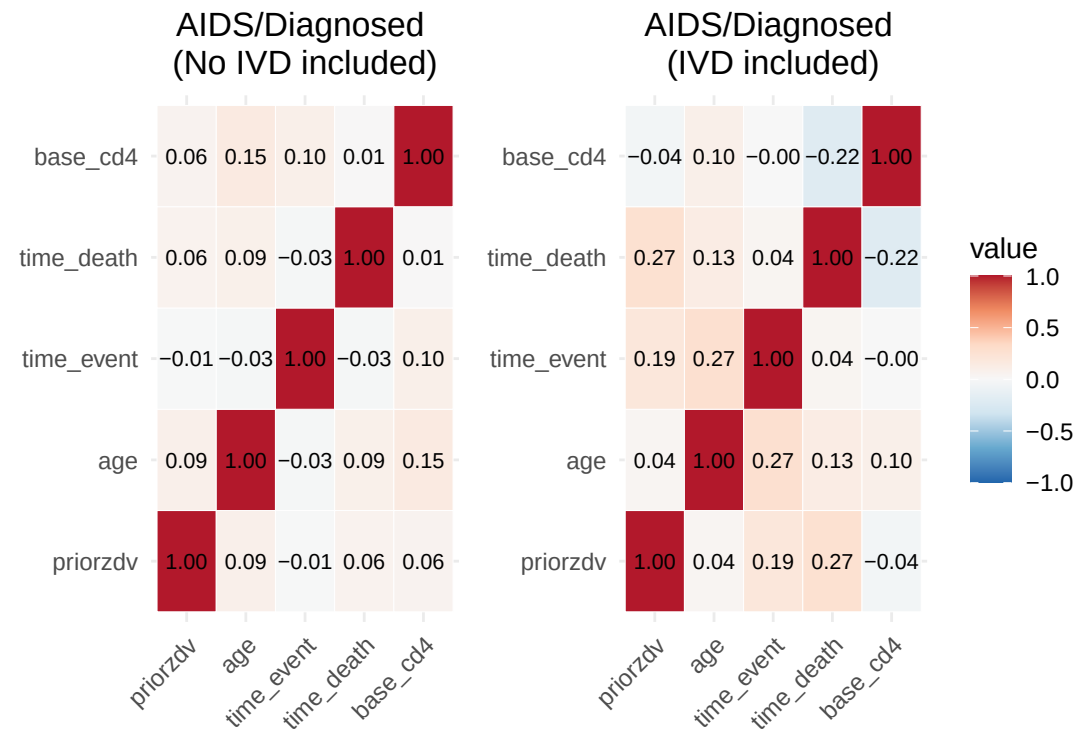
Correlations based on tx (1 = IDV included, 0 = otherwise) and outcome (AIDS/Death/NoAIDSNoDeath)

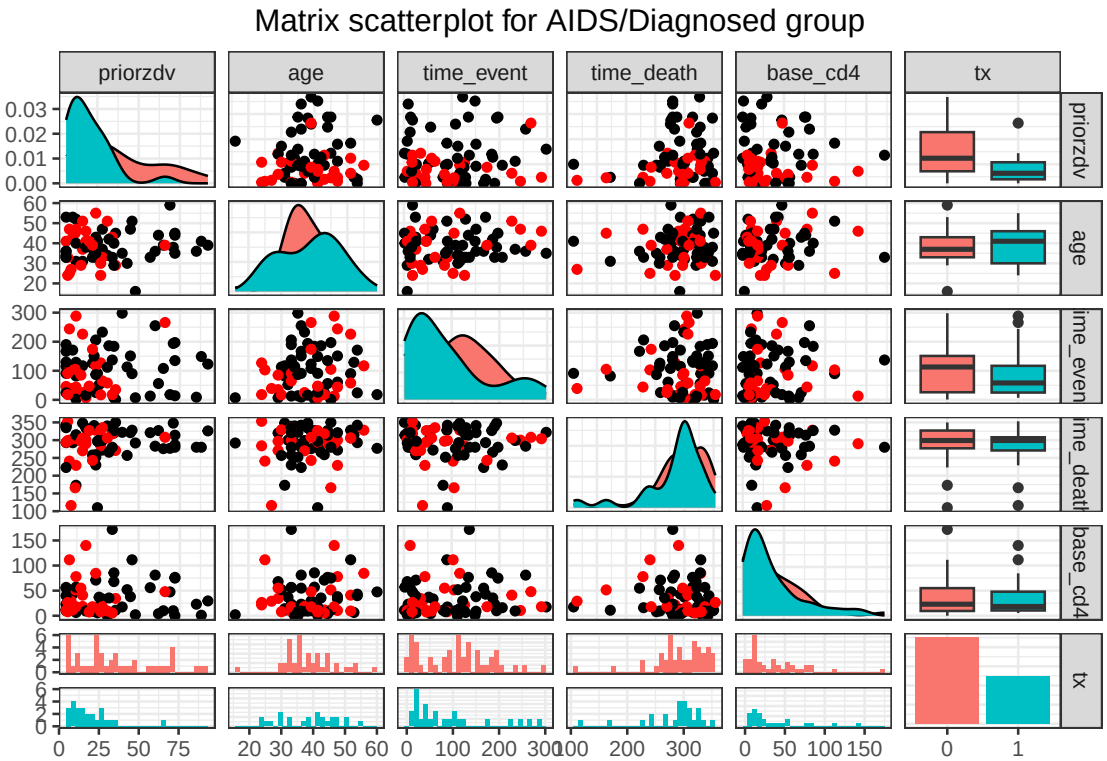
```
## [1] "NoDeath/NoDiagnosed"
```



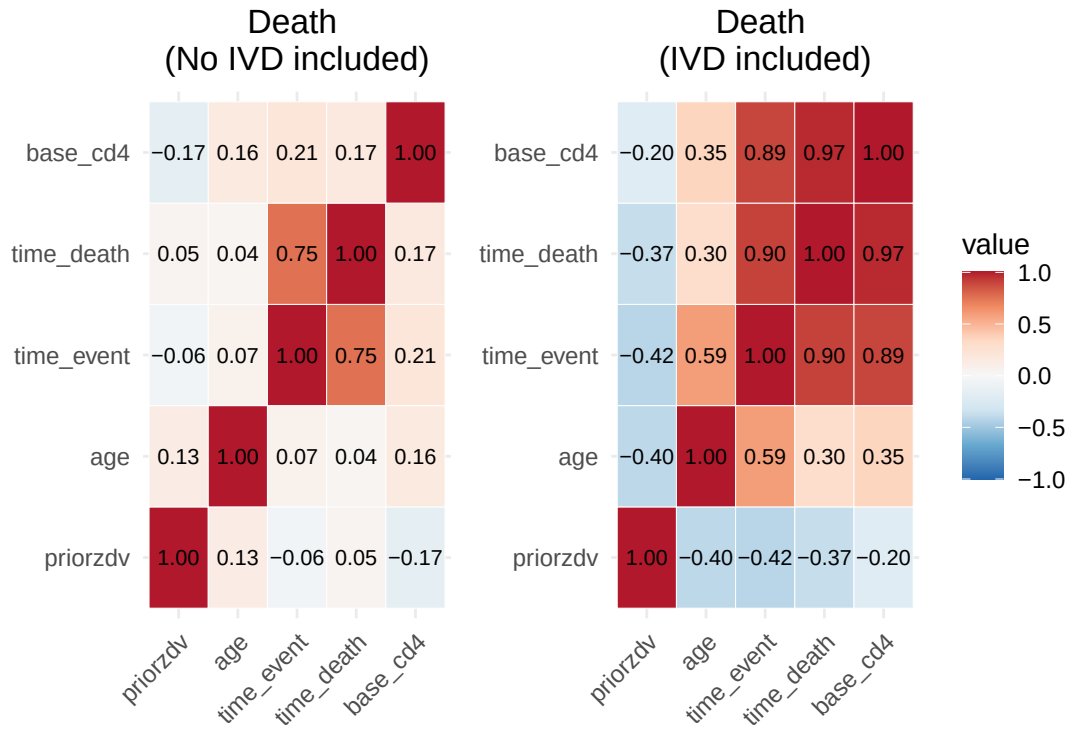


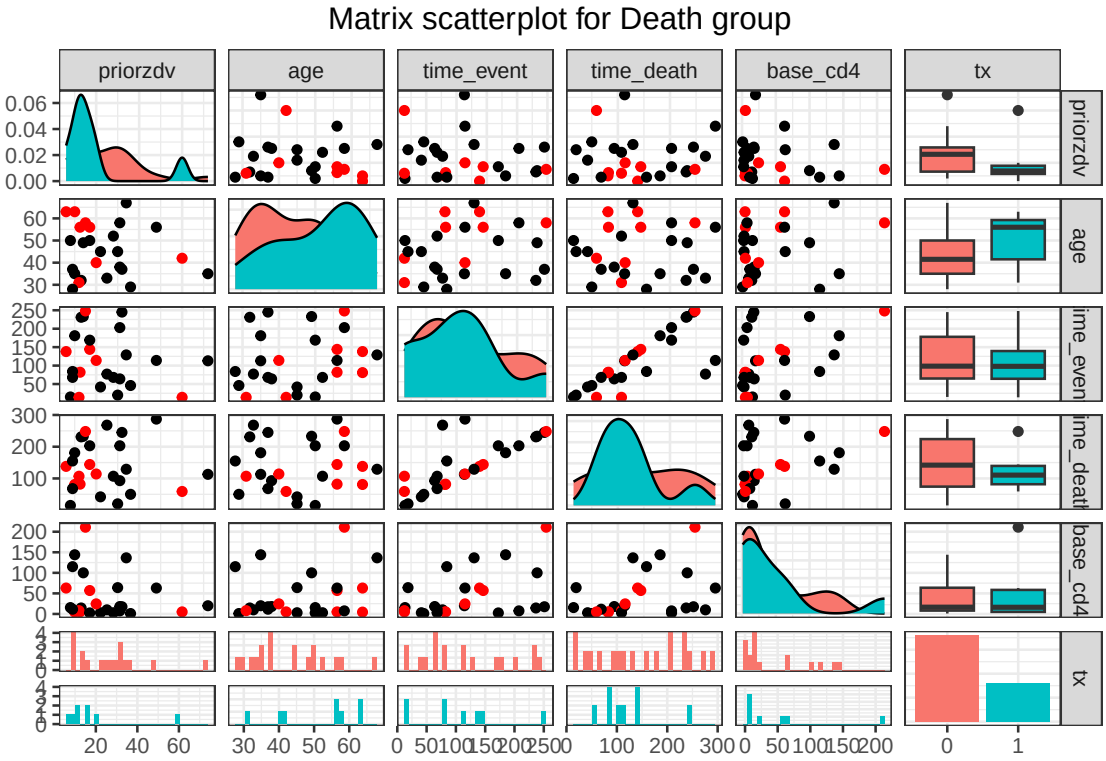
```
## [1] "AIDS/Diagnosed"
```






```
## [1] "Death"
```

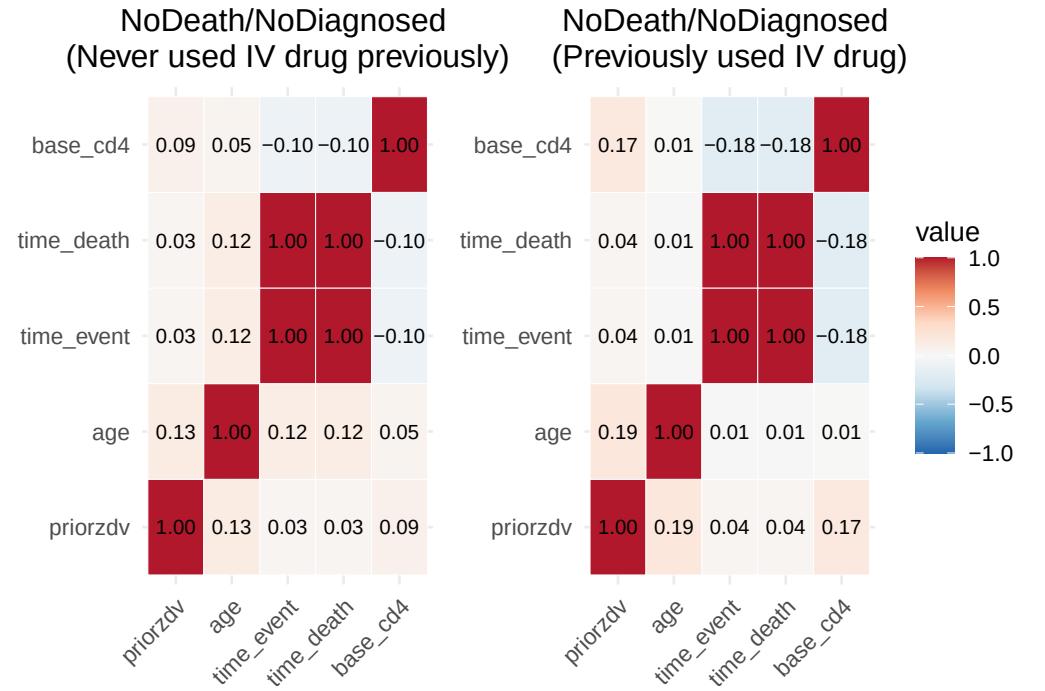


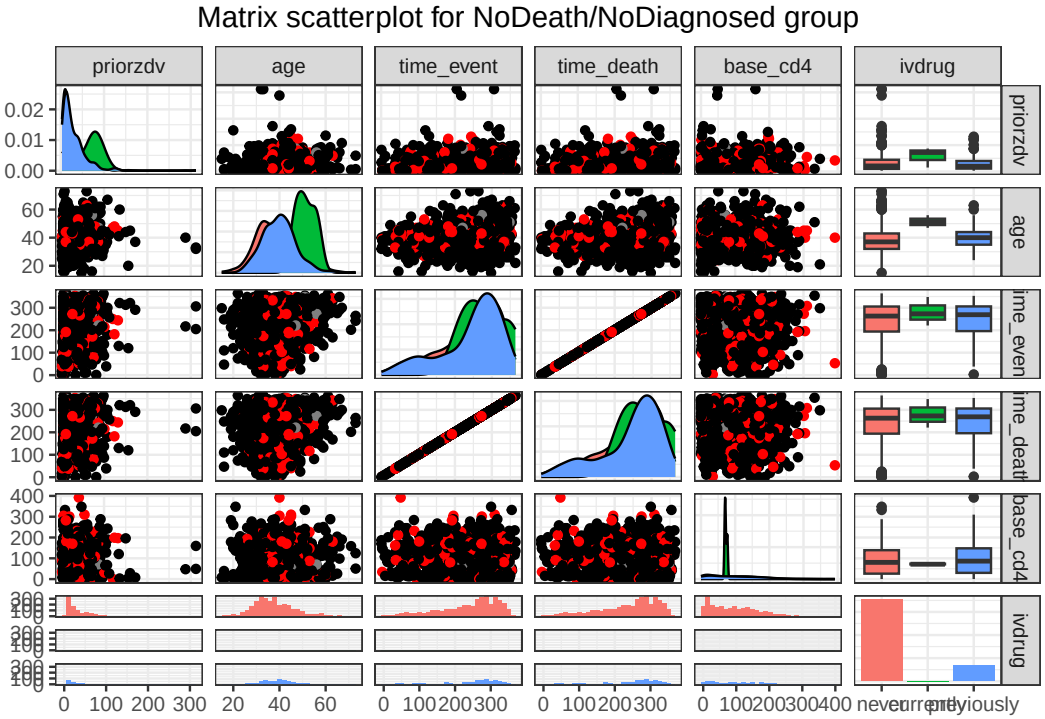


	AIDS/Diagnosed	Death	NoDeath/NoDiagnosed
0	45	18	514
1	25	8	541

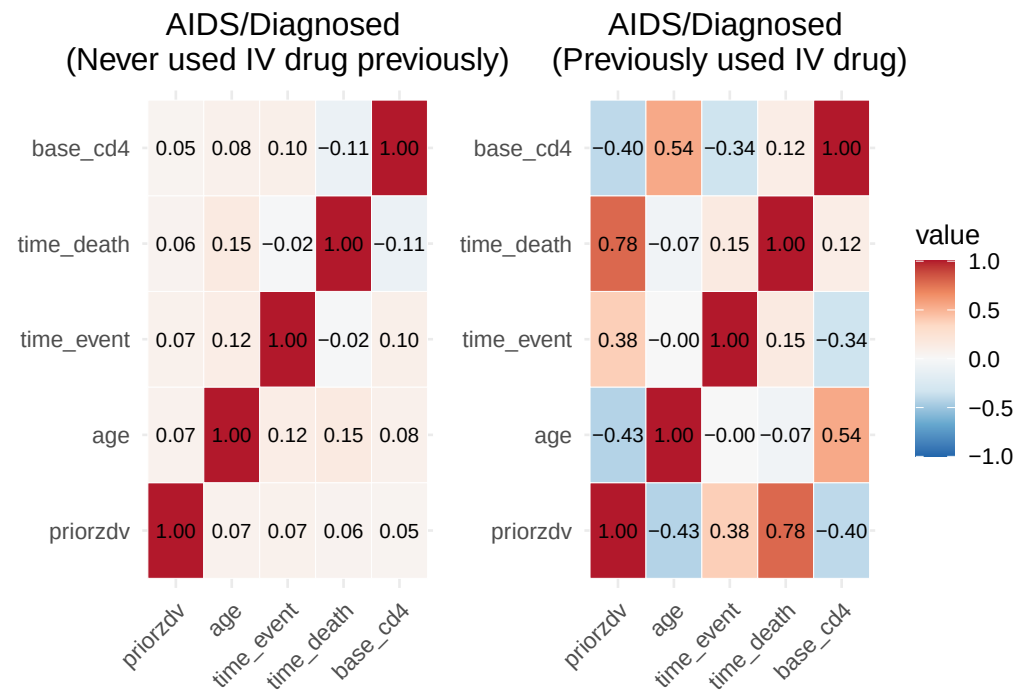
Correlations based on ivdrug (1 = IDV included, 0 = otherwise) and outcome (AIDS/Death/NoAIDSNoDeath

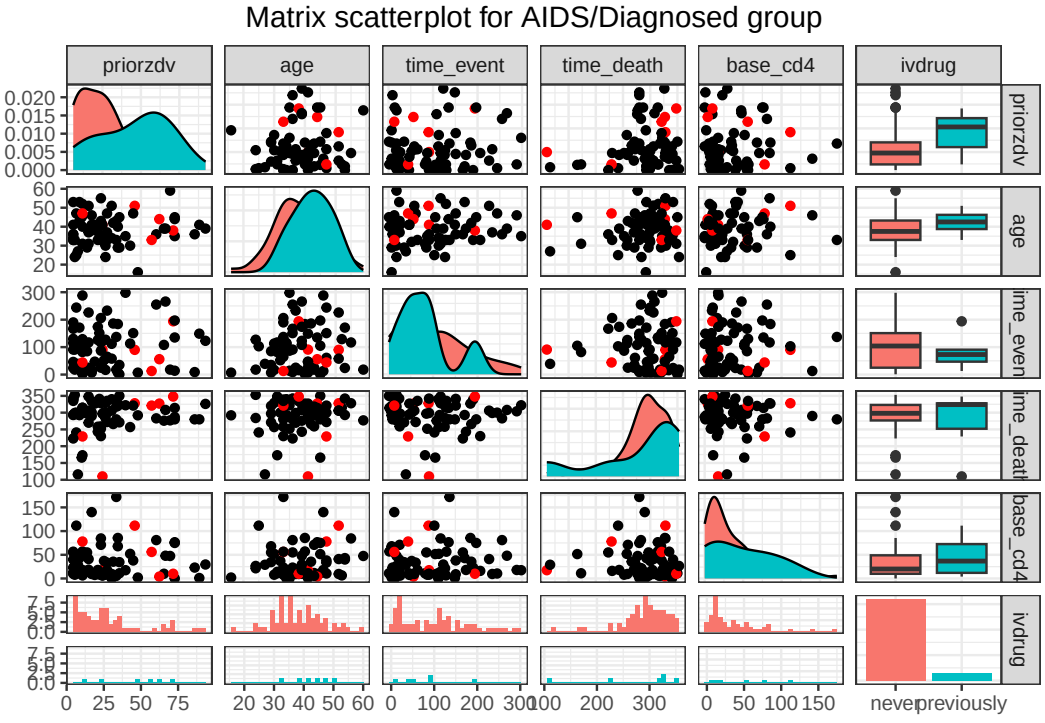
```
## [1] "NoDeath/NoDiagnosed"
```



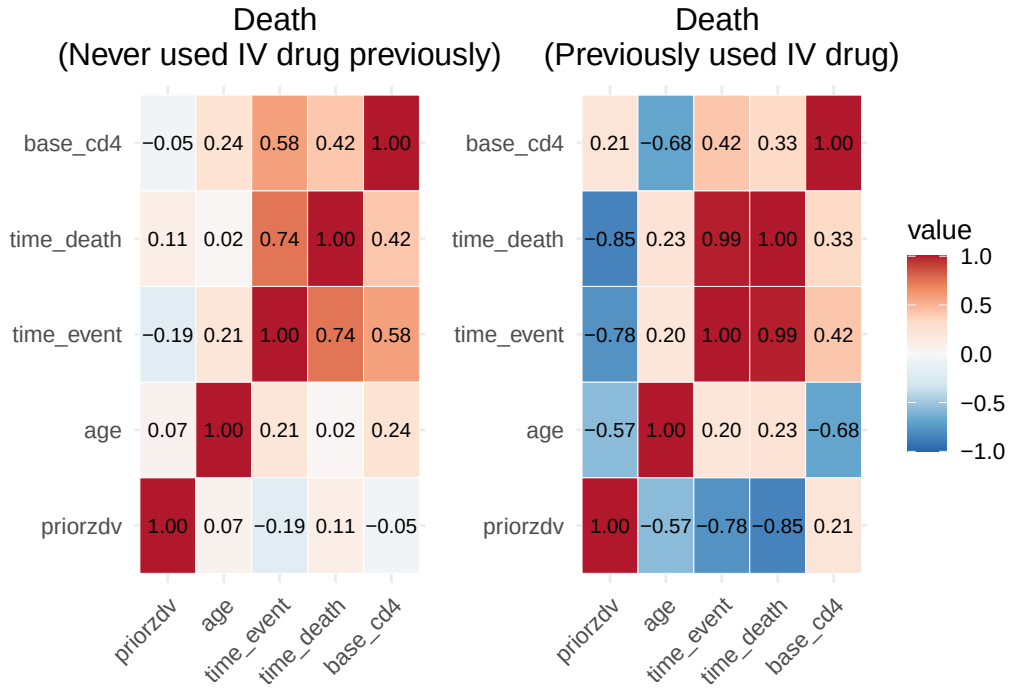


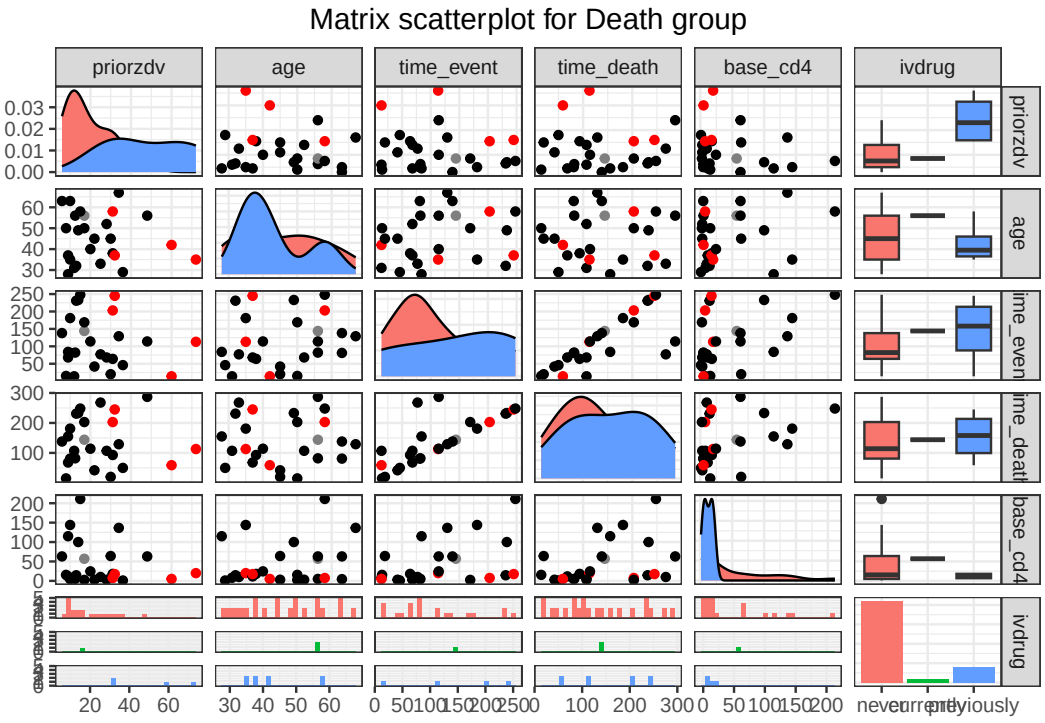

```
## [1] "AIDS/Diagnosed"
```


AIDS/Diagnosed
(Never used IV drug previously)
AIDS/Diagnosed
(Previously used IV drug)
base_cd4
time_death
time_event
age
priorzdv
0.05 0.08 0.10 −0.11 1.00
0.06 0.15 −0.02 1.00 −0.11
0.07 0.12 1.00 −0.02 0.10
0.07 1.00 0.12 0.15 0.08
1.00 0.07 0.07 0.06 0.05
−0.40 0.54 −0.34 0.12 1.00
0.78 −0.07 0.15 1.00 0.12
0.38 −0.00 1.00 0.15 −0.34
−0.43 1.00 −0.00 −0.07 0.54
1.00 −0.43 0.38 0.78 −0.40
priorzdv
age
time_event
time_death
base_cd4
value
1.0
0.5
0.0
−0.5
−1.0


Matrix scatterplot for AIDS/Diagnosed group
priorzdv
age
time_event
time_death
base_cd4
ivdrug
never
previously

```
## [1] "Death"
```

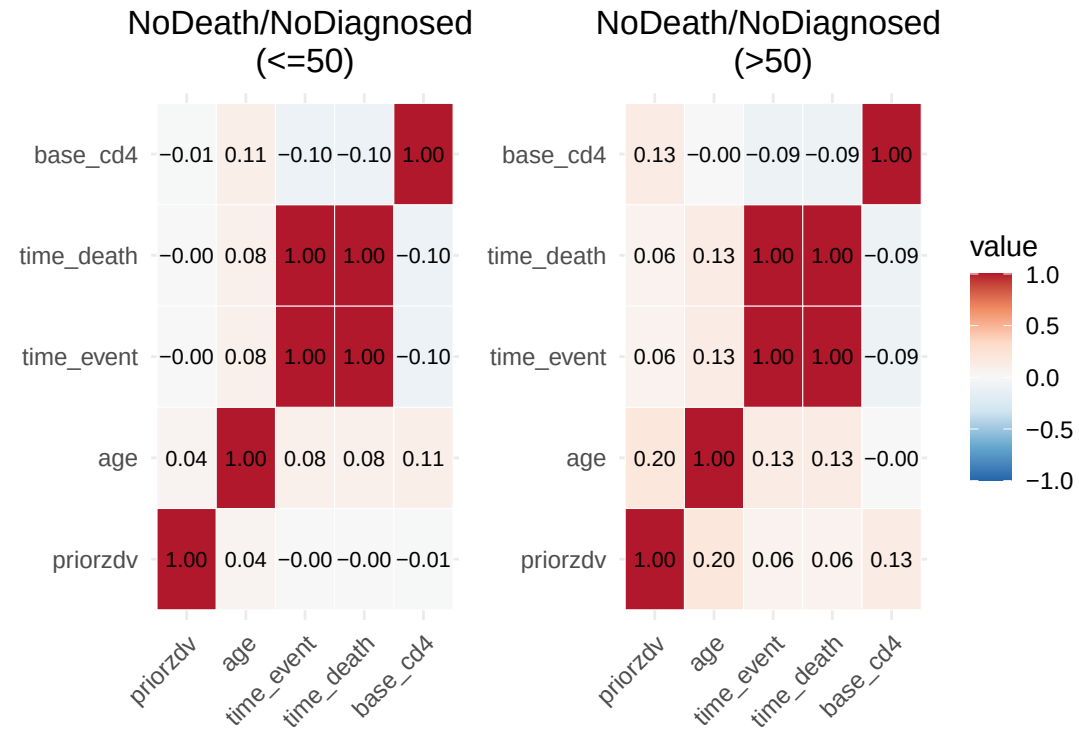


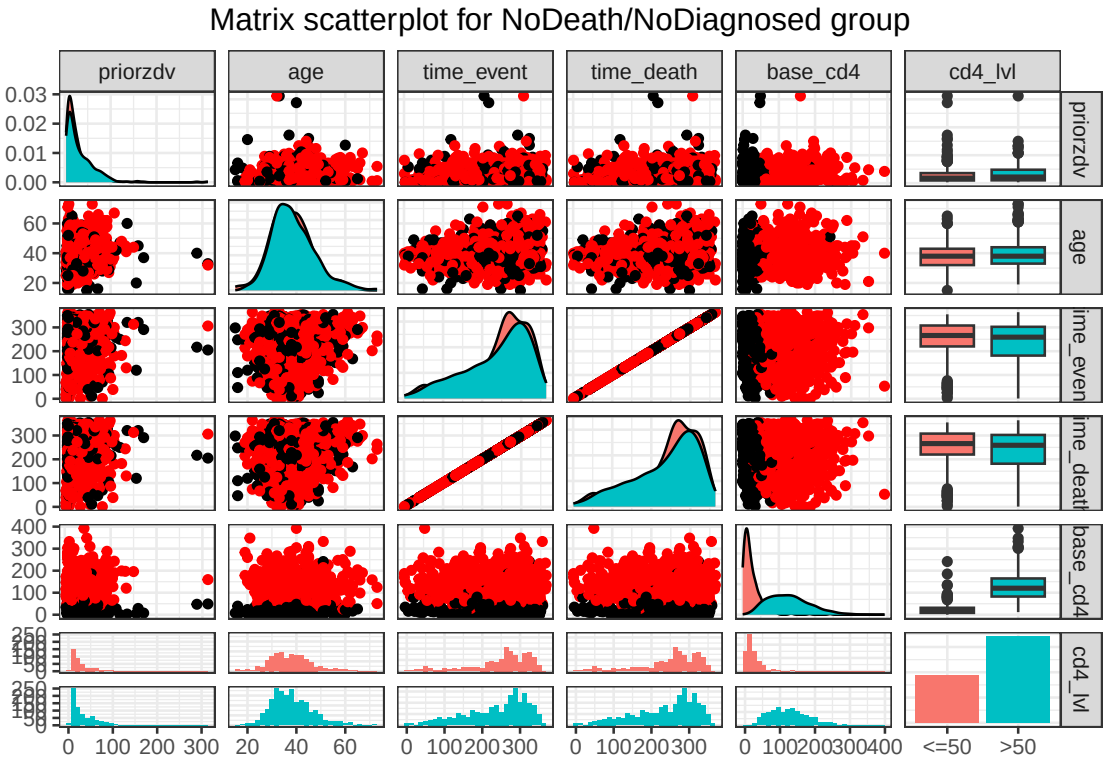


	AIDS/Diagnosed	Death	NoDeath/NoDiagnosed
never	64	21	883
currently	0	1	3
previously	6	4	169

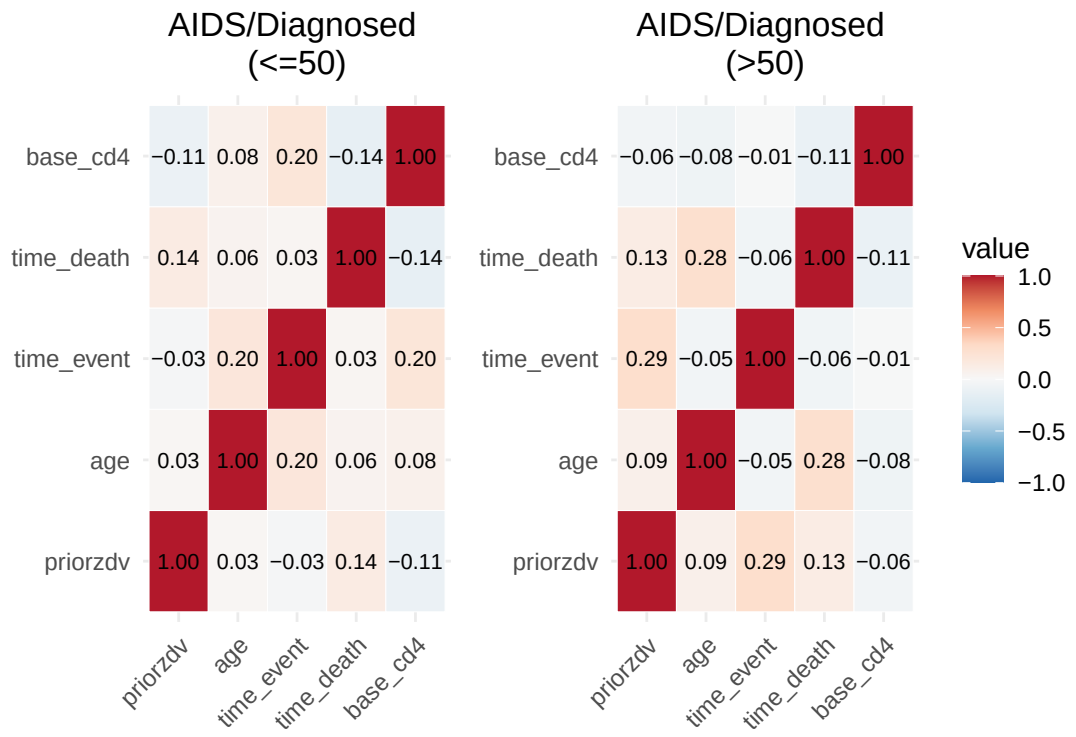
Correlations based on cd4_lvl (0 <= 50, 1 >50) and outcome (AIDS/Death/NoAIDSNoDeath)

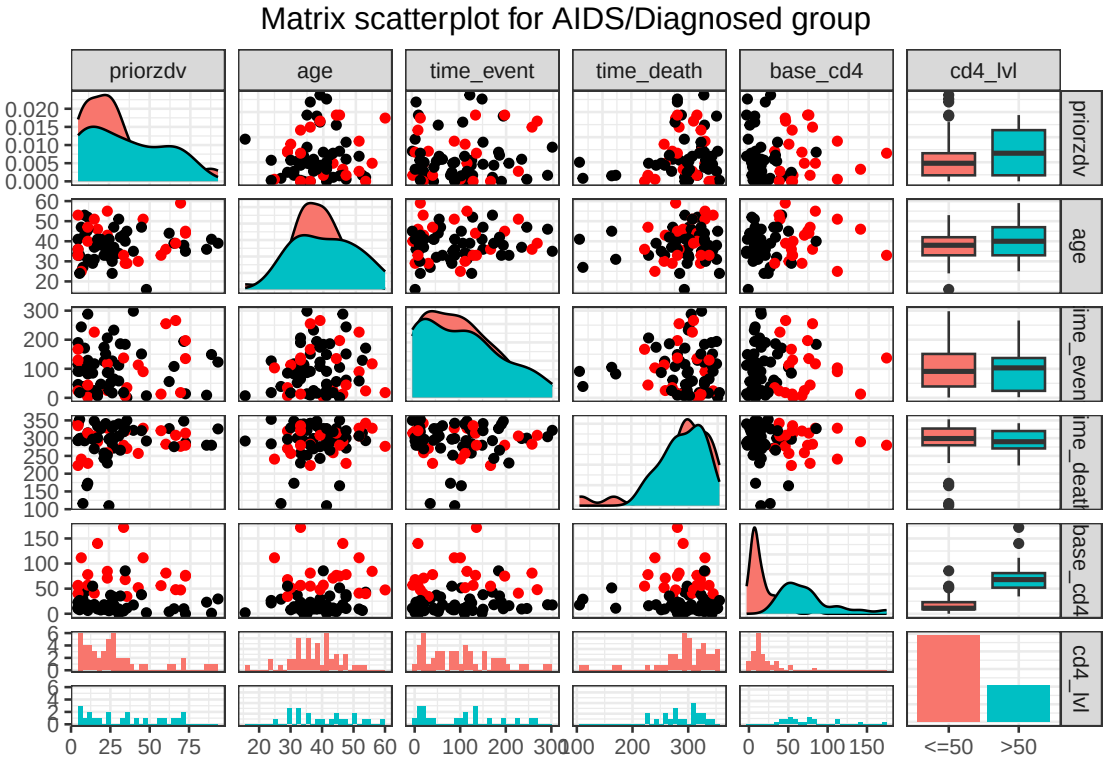
```
## [1] "NoDeath/NoDiagnosed"
```



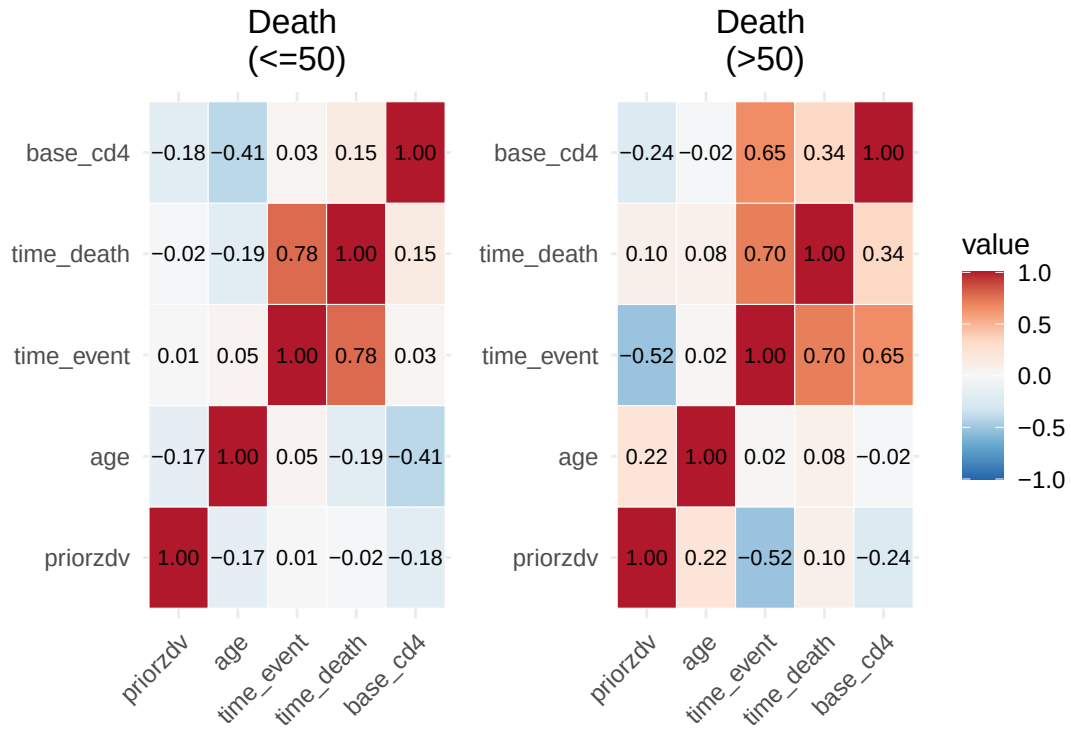


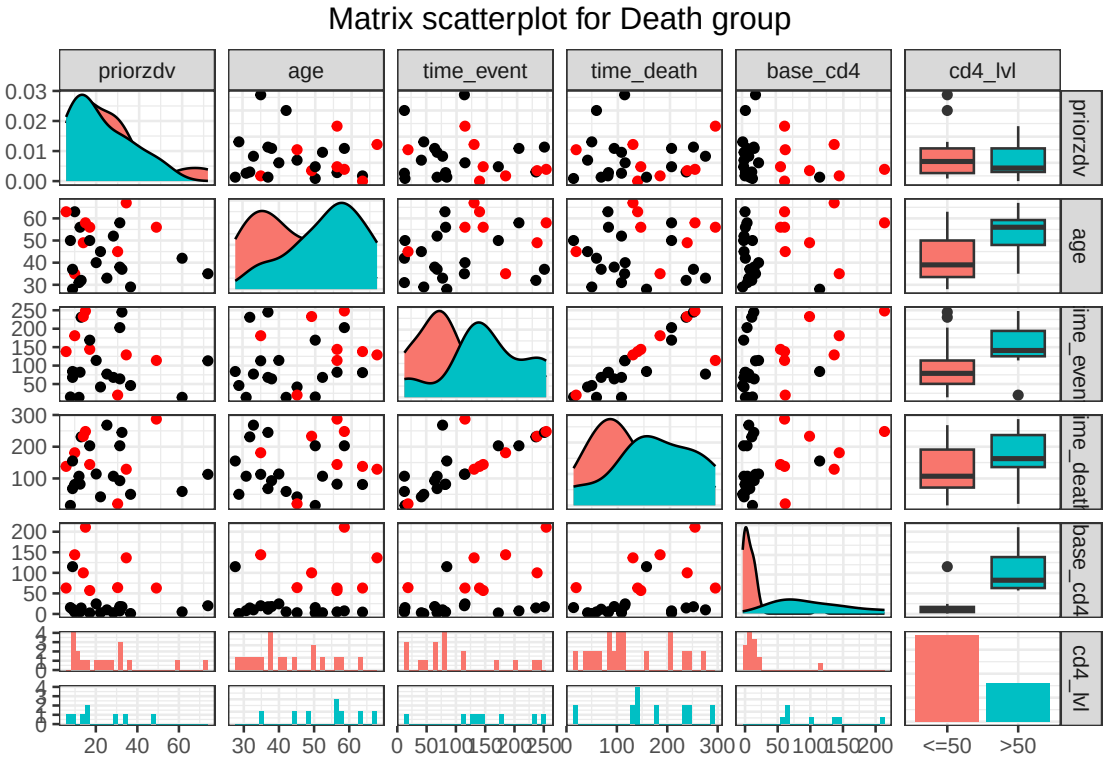
```
## [1] "AIDS/Diagnosed"
```






```
## [1] "Death"
```



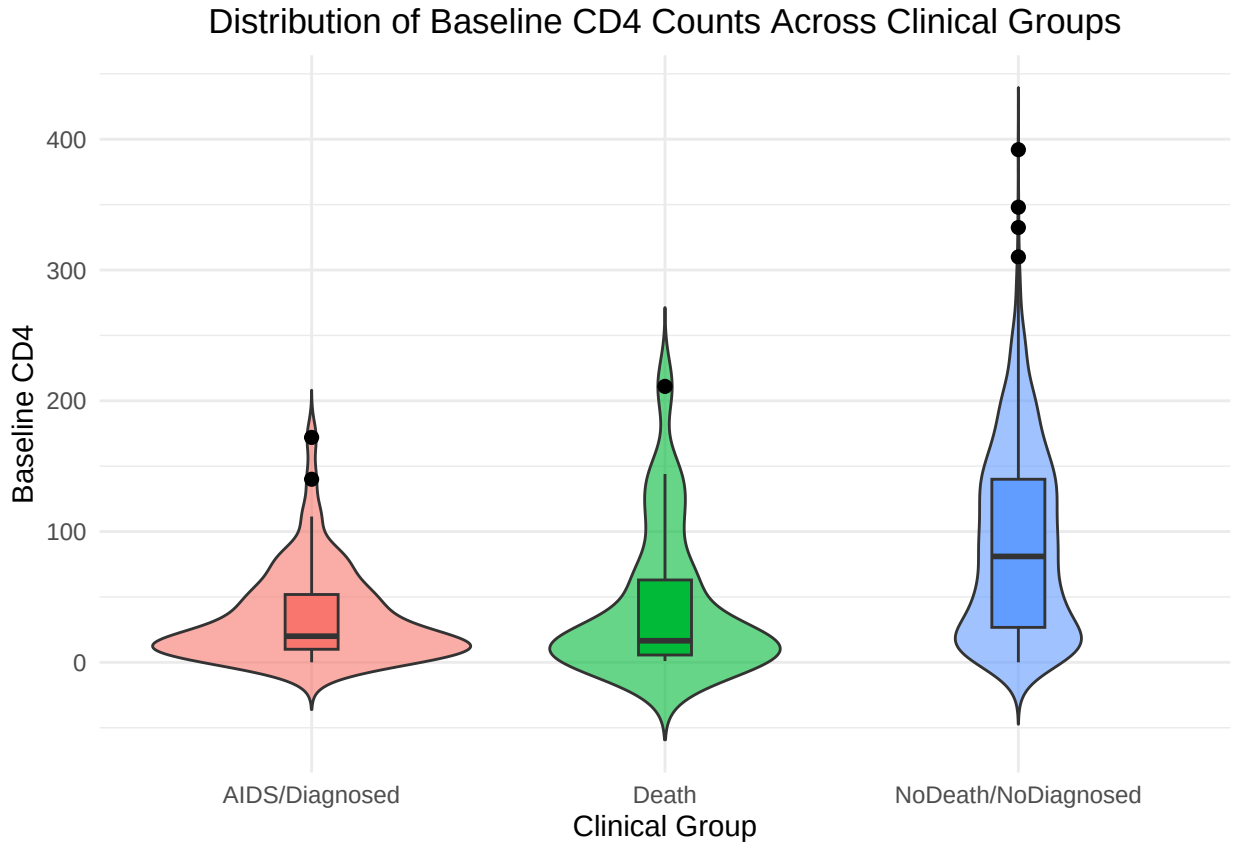


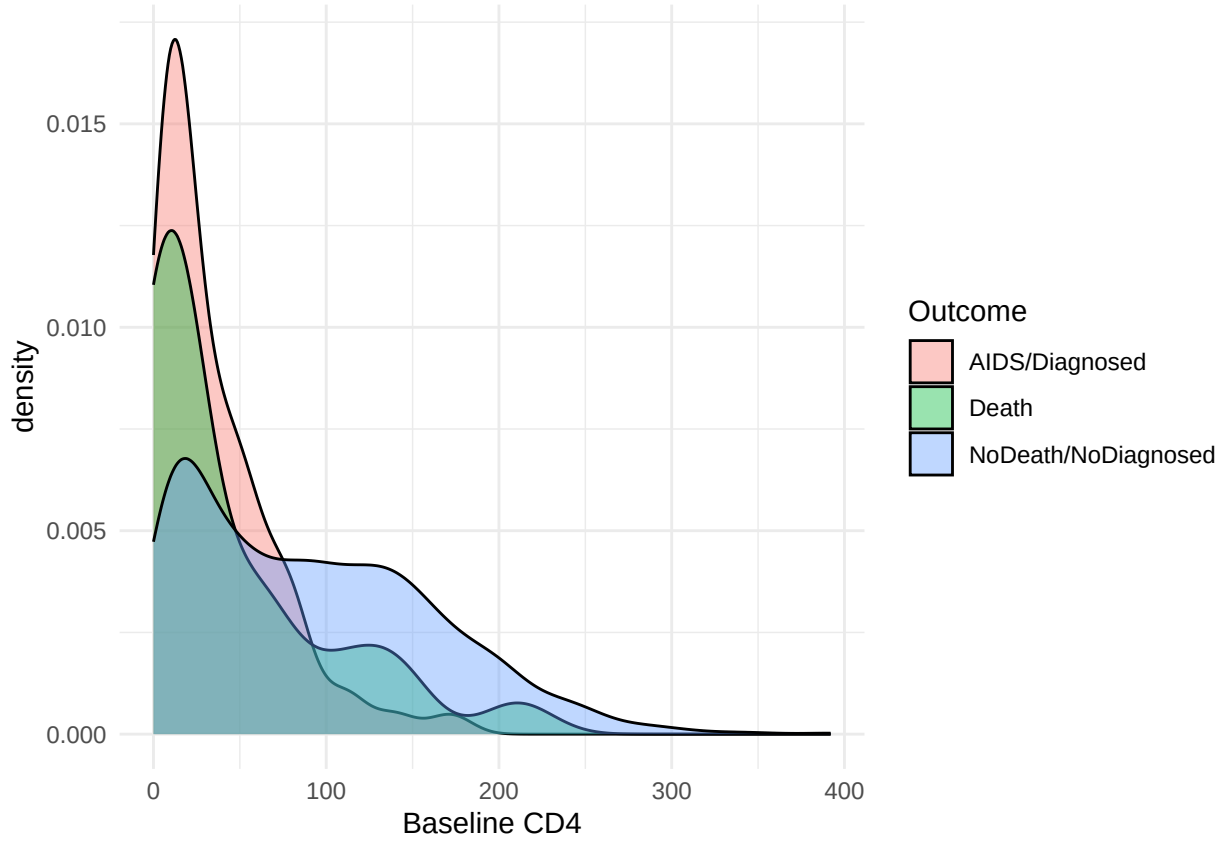
	AIDS/Diagnosed	Death	NoDeath/NoDiagnosed
<=50	49	18	372
>50	21	8	683

Understanding baseline_cd4 through other variables

outcome_diagnosed	n	mean_cd4	median_cd4	sd_cd4
AIDS/Diagnosed	70	34.8143	20.0	34.6426
Death	26	43.2500	16.5	55.2419
NoDeath/NoDiagnosed	1055	90.9513	81.0	70.4291

The violin plots illustrate a clear separation of baseline CD4 counts across clinical outcome groups: participants who developed AIDS or died had substantially lower CD4 levels at enrollment, whereas event-free participants exhibited higher and broader CD4 distributions. This visual pattern is confirmed numerically. Median baseline CD4 was 20 cells/mm³ in the AIDS/Diagnosed group and 16.5 cells/mm³ among those who died, compared with 81 cells/mm³ in the NoDeath/NoDiagnosed group. The interquartile ranges similarly reflect this gradient of immunological severity. Together, these results provide strong internal validation: baseline CD4 at enrollment effectively discriminates clinical trajectories, with lower immune reserves associated with subsequent AIDS-defining events or death.





Baseline CD4 counts showed a clear gradient across Karnofsky performance scores. Participants with poorer functional status (Karnofsky 70 and 80) exhibited markedly lower CD4 levels at enrollment, with median values of 22 and 39 cells/mm^3, respectively. In contrast, individuals with higher performance scores (90 and 100) had substantially higher CD4 medians (79 and 88 cells/mm^3). The widening distribution of CD4 at higher Karnofsky levels indicates greater immunological heterogeneity among clinically well participants. These findings are clinically consistent: better functional capacity reflects higher immune reserve, whereas severe immunosuppression is associated with diminished physical functioning.

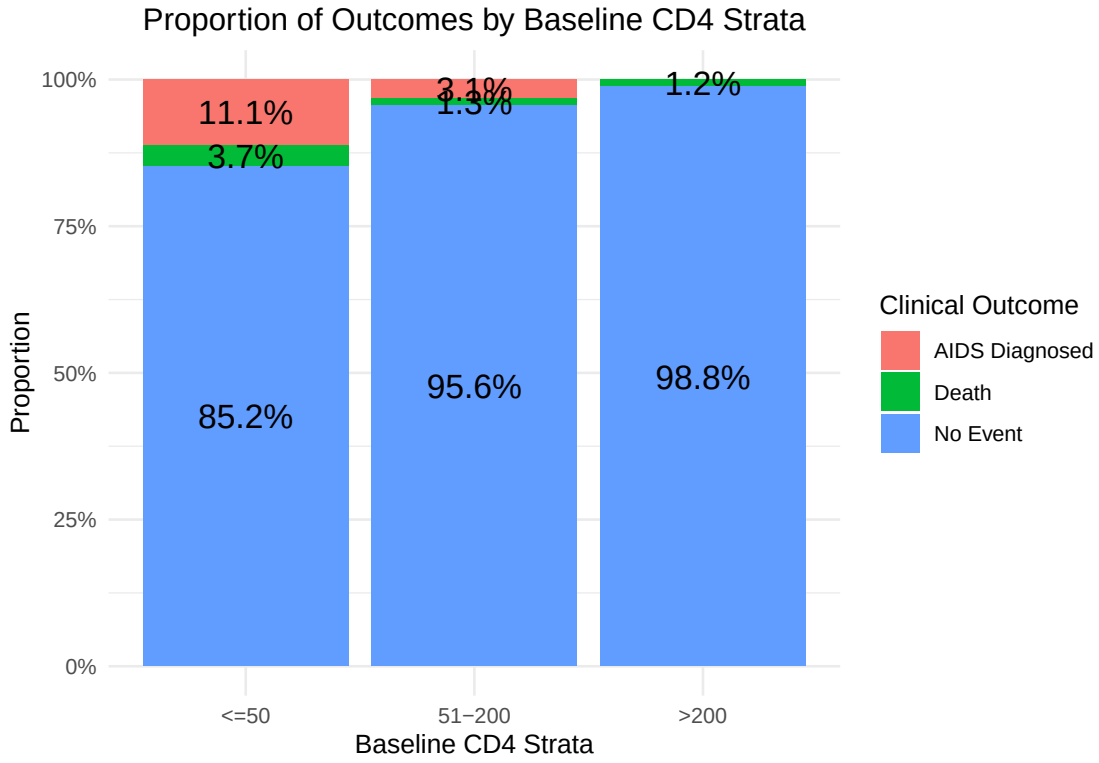
Considering those who had a base_cd4 higher than 200

Baseline immune status was severely compromised in this cohort: 88% of participants had CD4 less or equal than 200 at enrollment, and 40% had CD4 less or equal than 50. Such distributions are fully consistent with an advanced HIV-infected population and align with the high rates of AIDS-defining events and deaths observed during follow-up.

```
##
##   <=50 51-200   >200
##    459    610     82
```

	AIDS/Diagnosed	Death	NoDeath/NoDiagnosed
<=50	51	17	391
51–200	19	8	583
>200	0	1	81

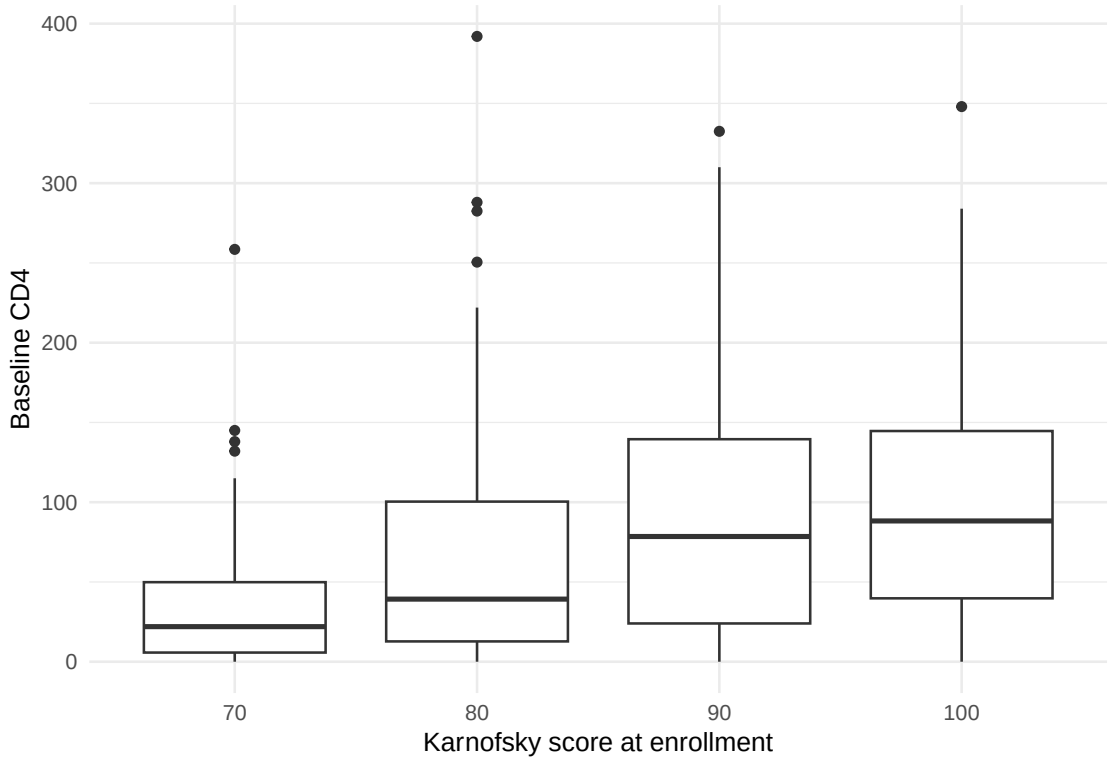
Stratifying participants by baseline CD4 count reveals a clear immunological gradient in clinical outcomes. Among individuals with CD4 ≤ 50 cells/mm^3, 15% experienced AIDS-defining events or death (51 diagnosed, 17 deaths), reflecting severe immunosuppression and high vulnerability. Participants with CD4 between 51–200 cells/mm^3 showed substantially lower event rates, while those with CD4 >200 exhibited almost no progression: none developed AIDS and only one death was recorded. These distributions confirm the expected biological pattern: lower immune reserves at enrollment are strongly associated with increased clinical deterioration during follow-up.



Considering the Karnofsky score

Baseline CD4 counts showed a clear gradient across Karnofsky performance scores. Participants with poorer functional status (Karnofsky 70 and 80) exhibited markedly lower CD4 levels at enrollment, with median values of 22 and 39 cells/mm^3, respectively. In contrast, individuals with higher performance scores (90 and 100) had substantially higher CD4 medians (79 and 88 cells/mm^3). The widening distribution of CD4 at higher Karnofsky levels indicates greater immunological heterogeneity among clinically well participants. These findings are clinically consistent: better functional capacity reflects higher immune reserve, whereas severe immunosuppression is associated with diminished physical functioning.

karnof	n	mean_cd4	median_cd4
70	32	43.6719	22.00
80	182	65.8434	39.25
90	541	88.9452	78.50
100	396	95.9971	88.25

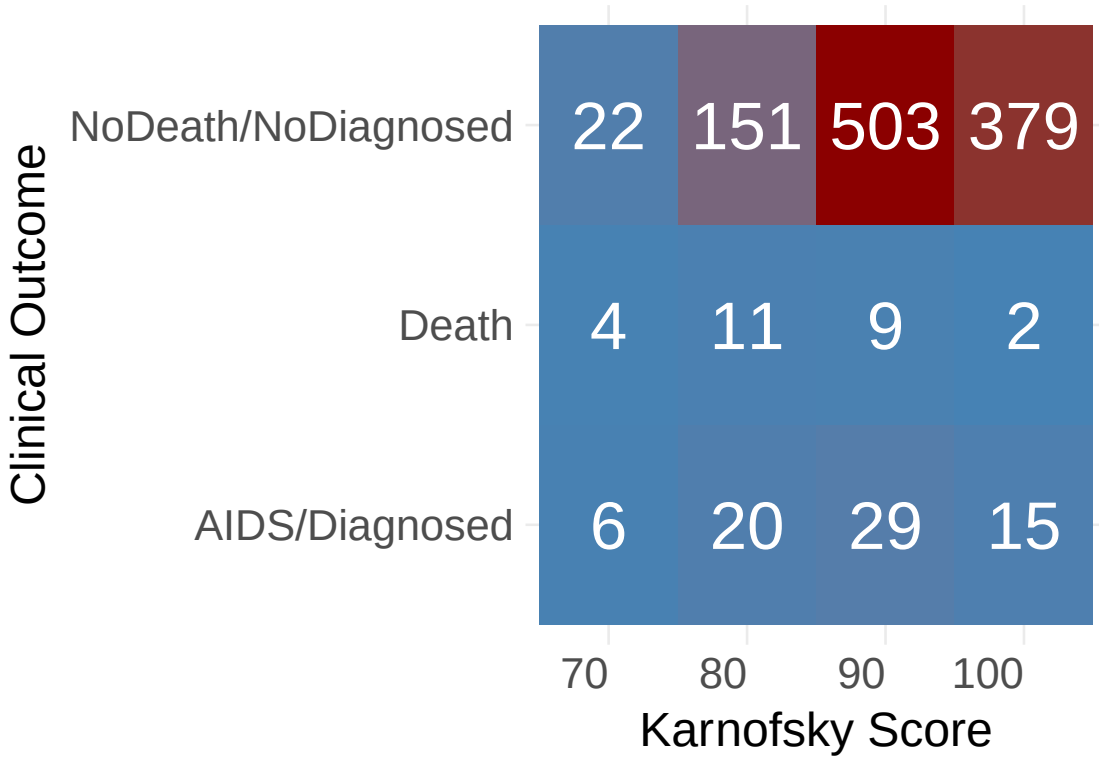


Across all analyses, Karnofsky performance status demonstrates a strong, consistent relationship with disease severity and clinical outcomes. Lower scores (70–80) cluster with low CD4 counts, higher rates of AIDS diagnosis, and significantly worse survival, whereas higher scores (90–100) align with preserved CD4 counts, minimal clinical events, and excellent survival trajectories. These findings highlight the Karnofsky index as a robust baseline prognostic marker in HIV-infected individuals.

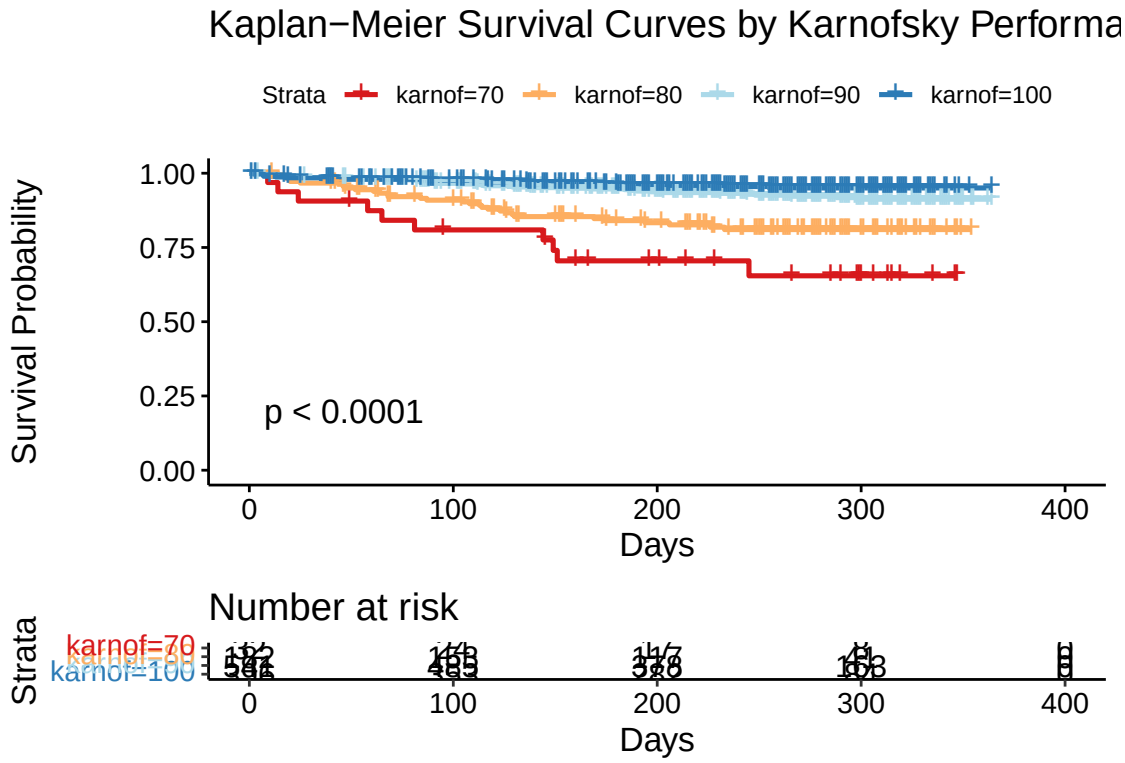
Higher Karnofsky performance scores were strongly associated with favorable clinical outcomes. Participants with scores of 90–100 exhibited very low rates of AIDS diagnosis and death, whereas those with scores of 70–80 had substantially higher event rates. This supports the clinical role of performance status as a global indicator of disease severity and short-term prognosis.

Table 10: Proportions by Karnofsky score of functional status

	AIDS/Diagnosed	Death	NoDeath/NoDiagnosed
70	0.19	0.12	0.69
80	0.11	0.06	0.83
90	0.05	0.02	0.93
100	0.04	0.01	0.96

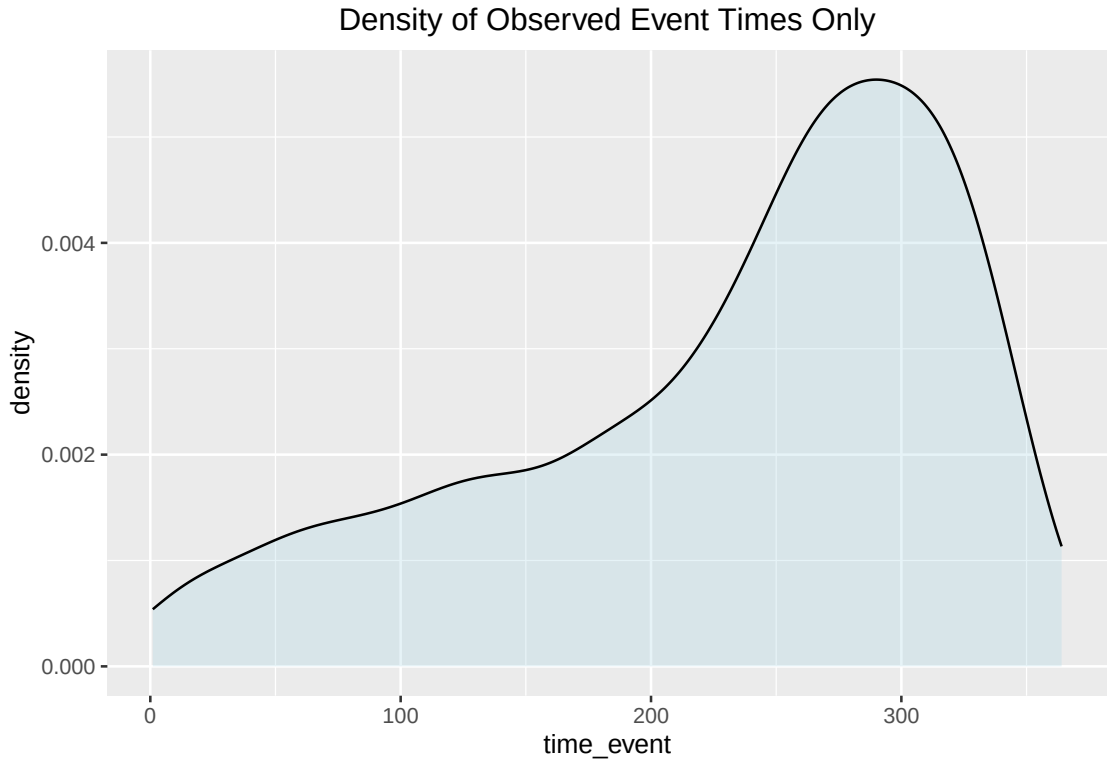


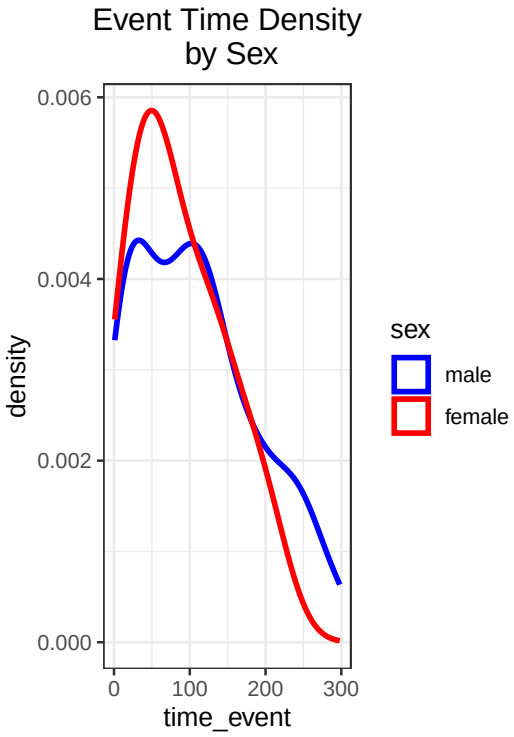
Survival differed substantially across Karnofsky performance categories. Participants with the lowest performance score (70) experienced markedly higher mortality within the first year, whereas those with scores of 90–100 had excellent survival with minimal drop-off. These patterns reinforce the Karnofsky score as a sensitive predictor of short-term mortality risk in HIV patients at baseline.

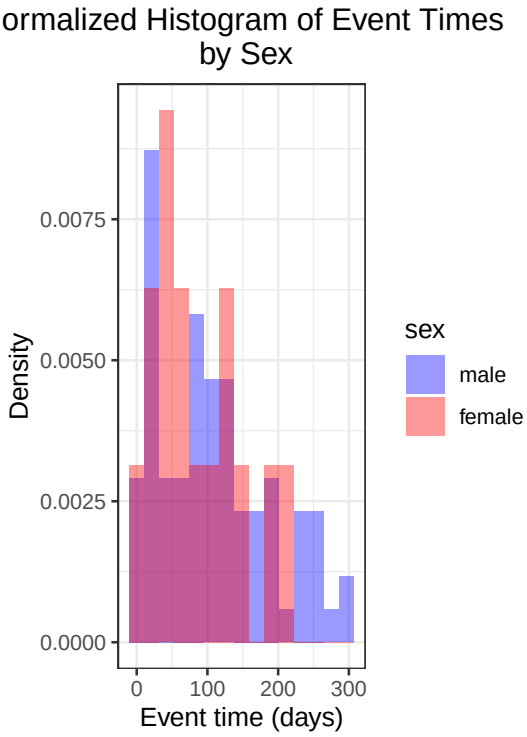


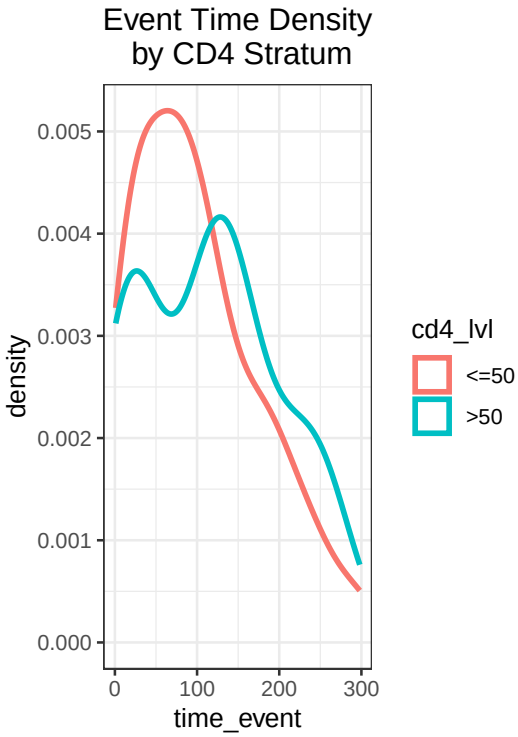
Follow-Up Pattern and Censoring Mechanism

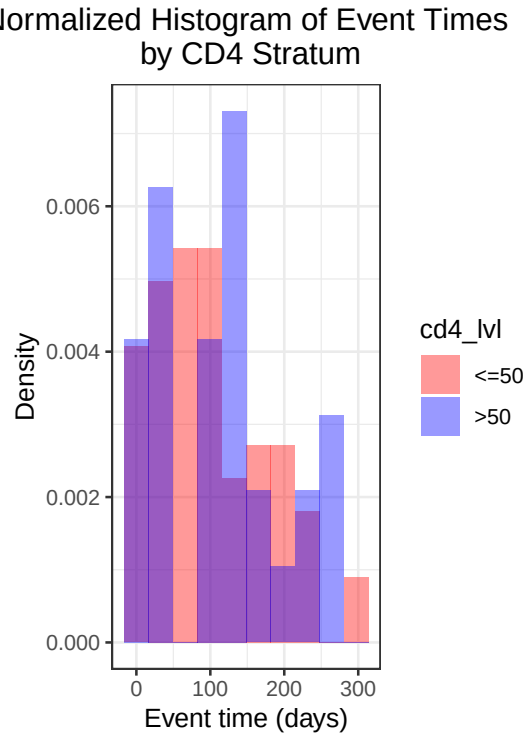
An examination of the event-time distributions, censoring-time distributions, and combined follow-up times indicates that the study is characterized by substantial censoring near the end of the follow-up window (approximately 250–360 days). The censoring densities are highly similar across all categorical covariates, suggesting **no differential censoring** by subgroup. The shape and timing of the censoring distribution—together with the fact that censoring predominantly occurs at a common study termination time—provide strong evidence that censoring is non-informative with respect to the event process. Moreover, the event-time densities peak earlier than the censoring densities, which is consistent with adequate follow-up rather than early loss to follow-up. Although some small subgroups (e.g., current IV drug users or minor race categories) exhibit irregular density shapes due to limited sample sizes, these patterns do not affect the overall conclusion. Taken together, the follow-up characteristics appear compatible with the **independent (non-informative) censoring assumption**, which is explicitly required for the Kaplan–Meier estimator, Nelson–Aalen estimator, Cox proportional hazards model, and log-rank tests, as discussed in the course lectures.

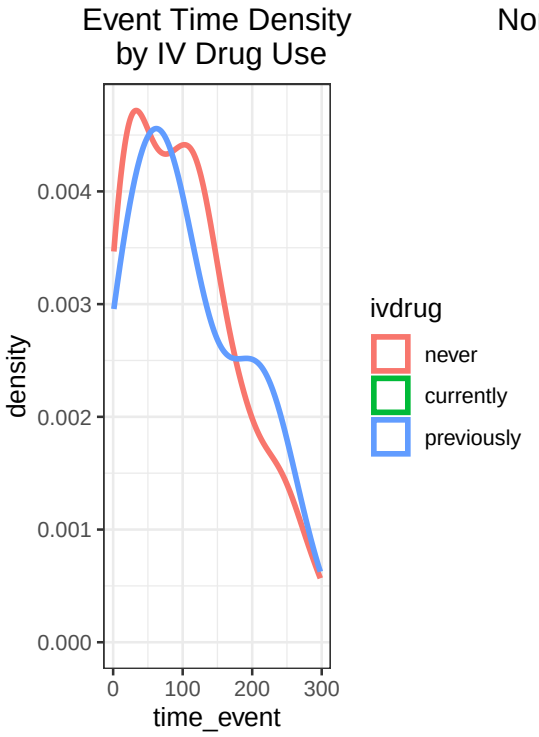


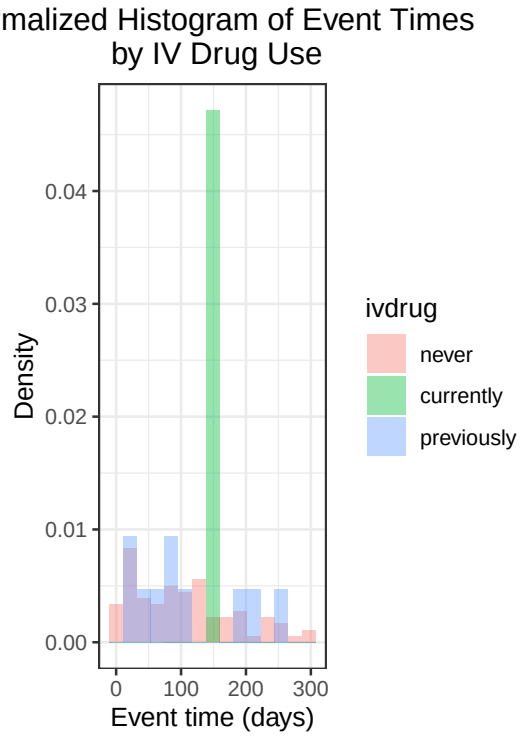

Event Time Density
by Sex
density
0.006
0.004
0.002
0.000
0
100
200
300
time_event
sex
male
female

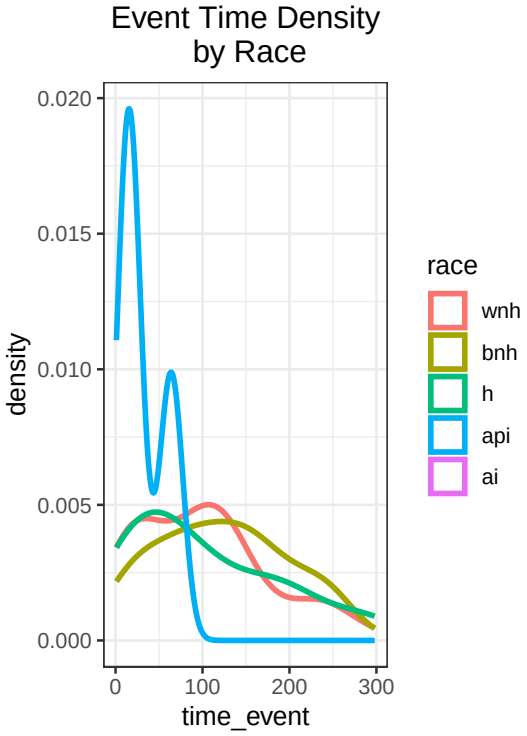

Normalized Histogram of Event Times
by Sex
Density
0.0075
0.0050
0.0025
0.0000
0
100
200
300
Event time (days)
sex
male
female

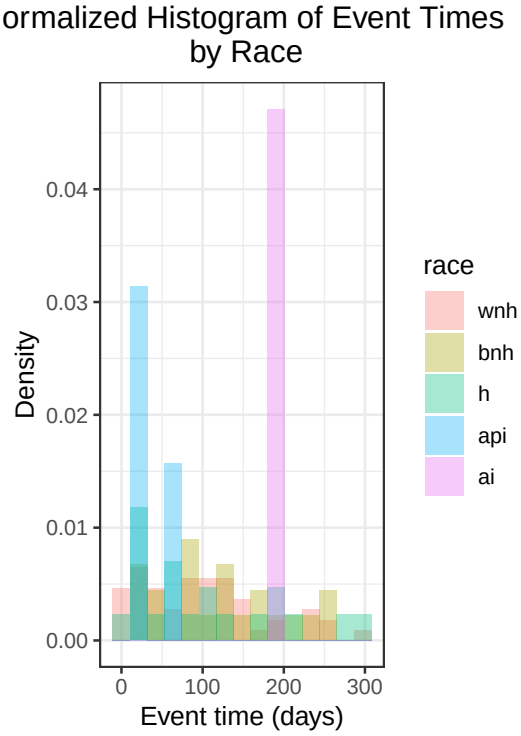

Event Time Density
by CD4 Stratum
density
0.005
0.004
0.003
0.002
0.001
0.000
0
100
200
300
time_event
cd4_lvl
<=50
>50

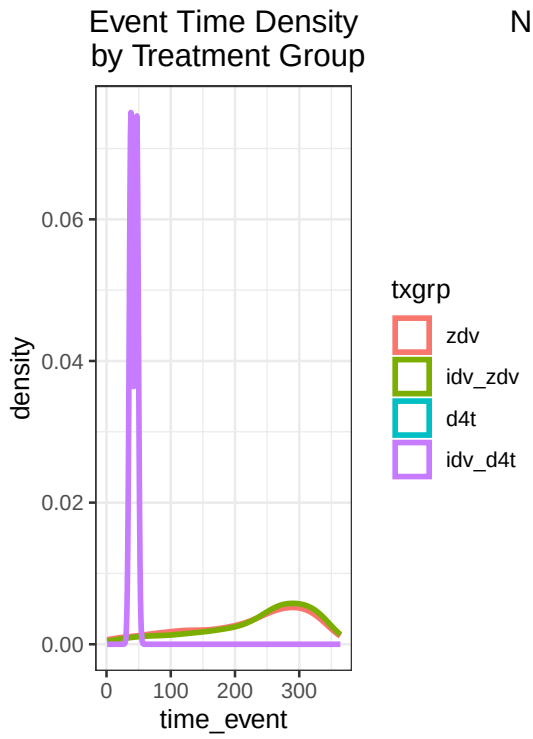

Normalized Histogram of Event Times
by CD4 Stratum
Density
0.006
0.004
0.002
0.000
0
100
200
300
Event time (days)
cd4_lvl
<=50
>50

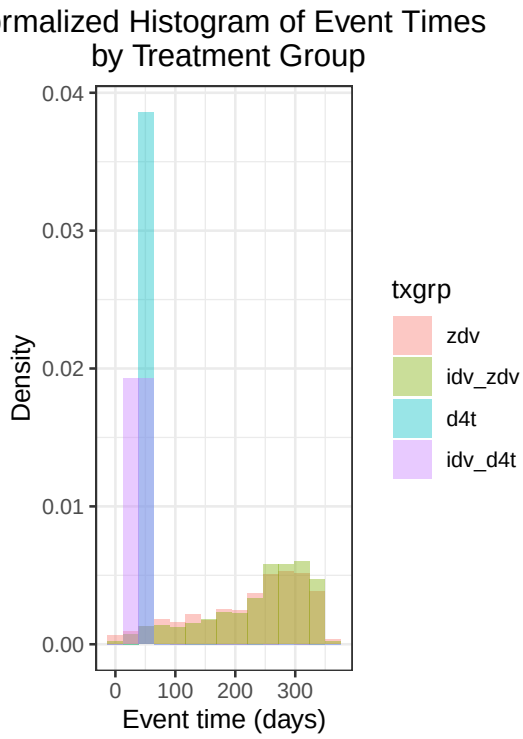

Event Time Density
by IV Drug Use
density
0.004
0.003
0.002
0.001
0.000
0
100
200
300
time_event
ivdrug
never
currently
previously

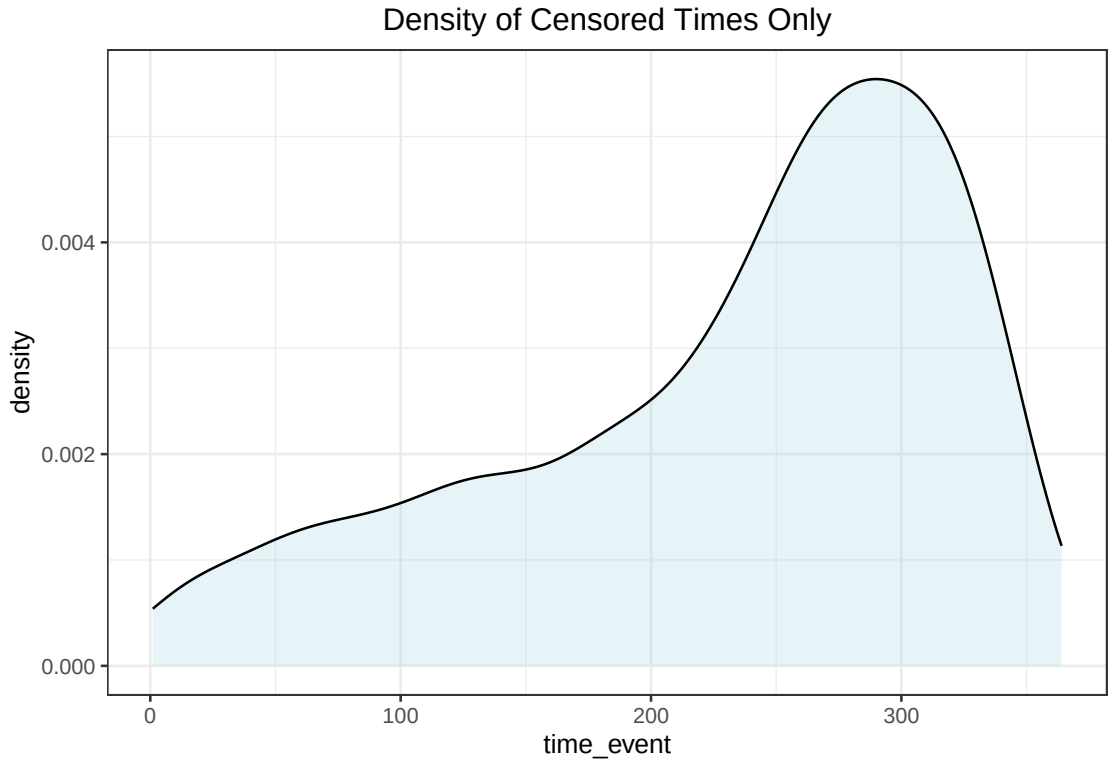

Normalized Histogram of Event Times
by IV Drug Use
Density
0.04
0.03
0.02
0.01
0.00
0
100
200
300
Event time (days)
ivdrug
never
currently
previously

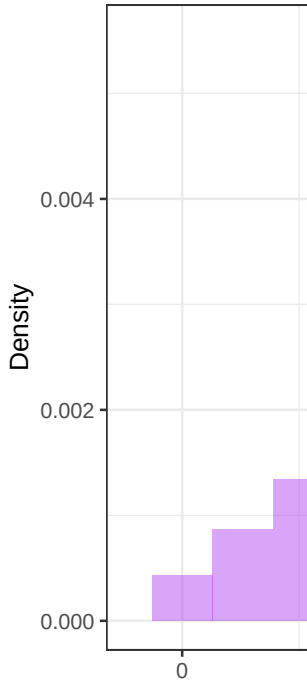

Event Time Density
by Race
density
0.020
0.015
0.010
0.005
0.000
0
100
200
300
time_event
race
wnh
bnh
h
api
ai

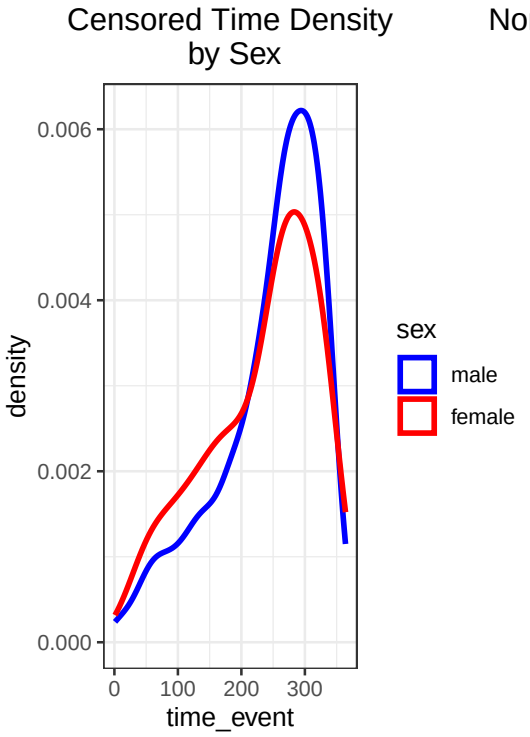

Normalized Histogram of Event Times
by Race
Density
0.04
0.03
0.02
0.01
0.00
0
100
200
300
Event time (days)
race
wnh
bnh
h
api
ai

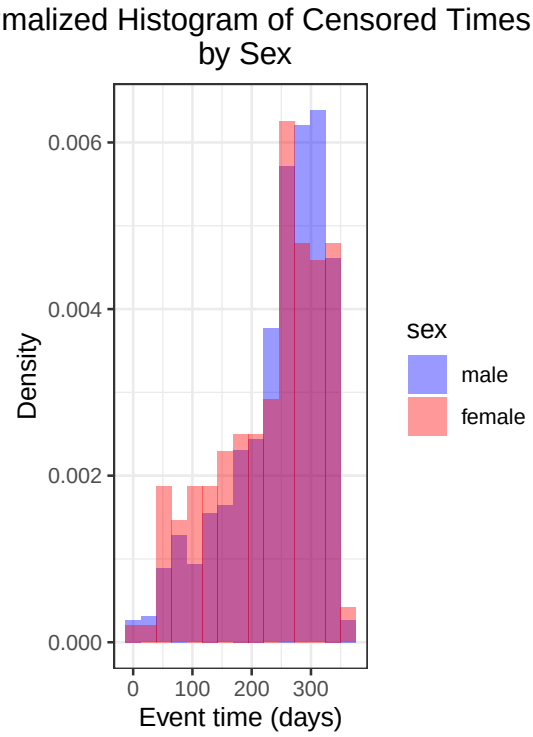

Event Time Density
by Treatment Group
density
0.06
0.04
0.02
0.00
0
100
200
300
time_event
txgrp
zdv
idv_zdv
d4t
idv_d4t

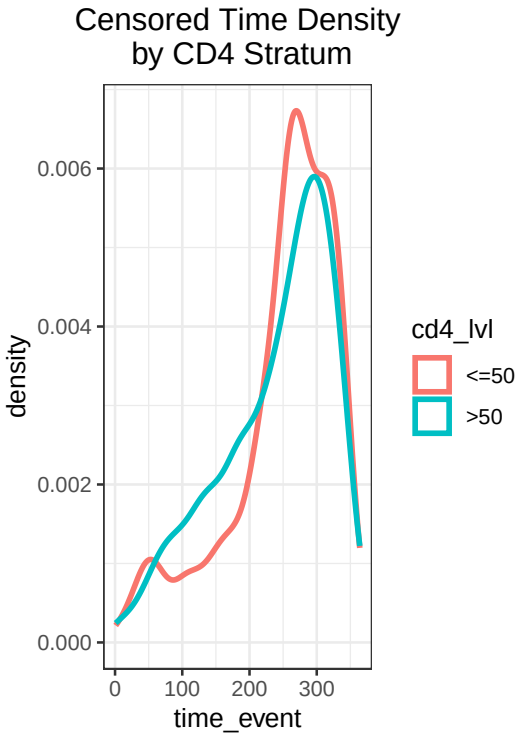

Normalized Histogram of Event Times
by Treatment Group
Density
0.04
0.03
0.02
0.01
0.00
0
100
200
300
Event time (days)
txgrp
zdv
idv_zdv
d4t
idv_d4t

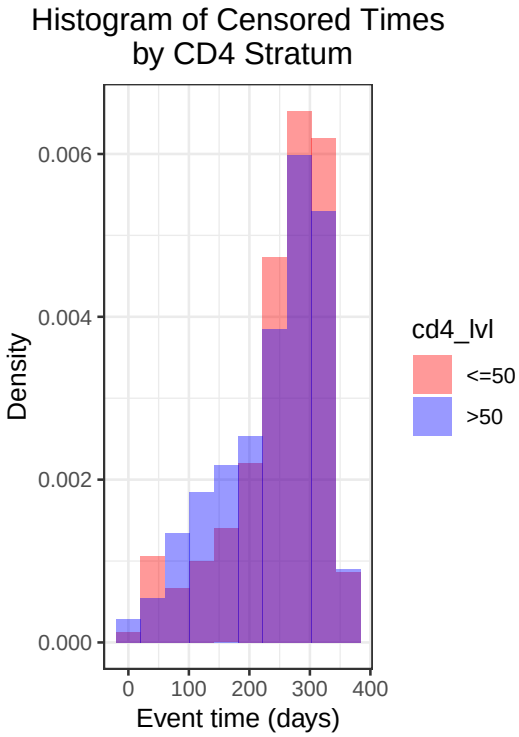

Density of Censored Times Only
density
0.004
0.002
0.000
0
100
200
300
time_event

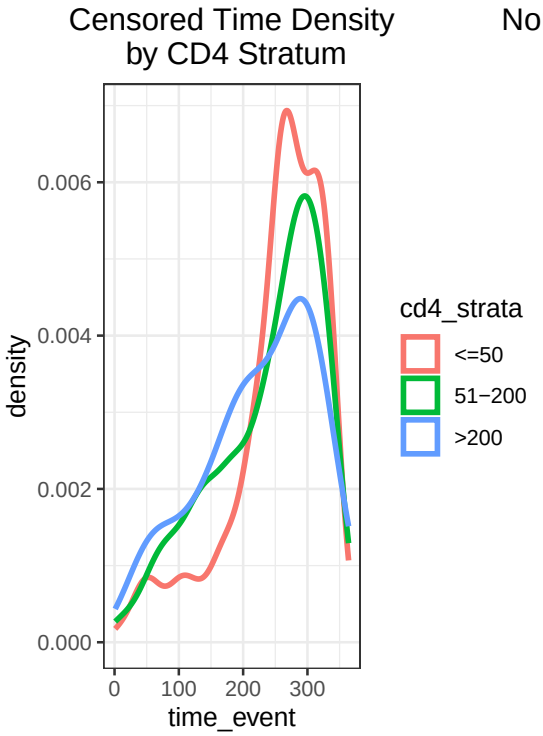

Density
0.004
0.002
0.000
0

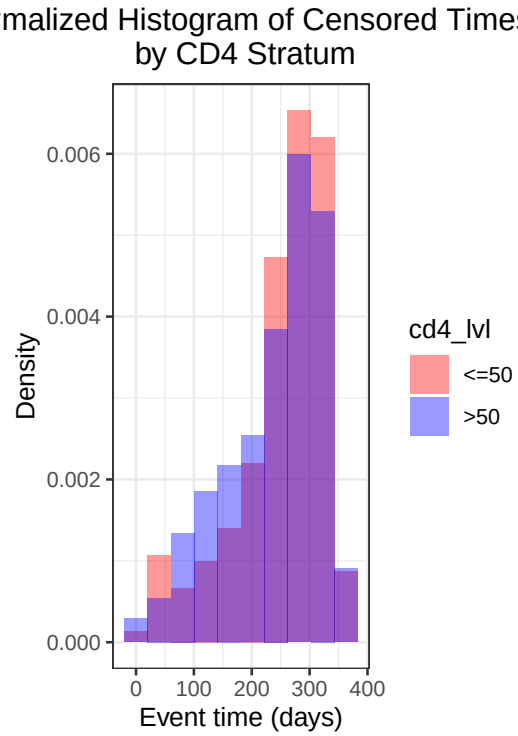

Censored Time Density
by Sex
density
0.006
0.004
0.002
0.000
0
100
200
300
time_event
sex
male
female

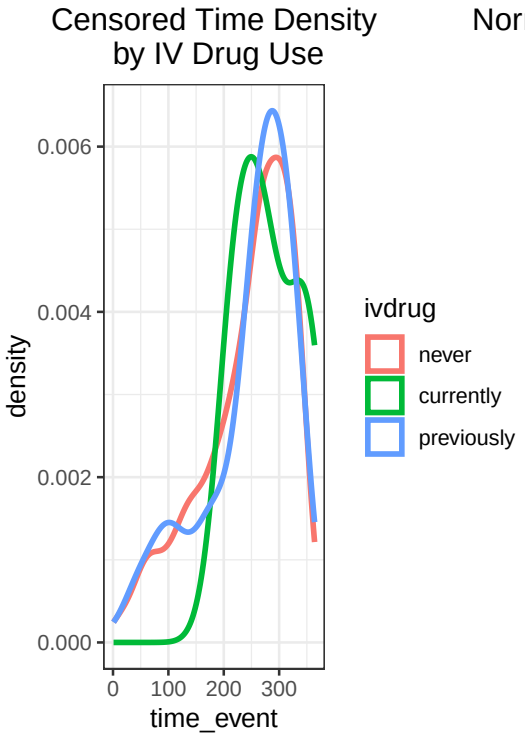

Normalized Histogram of Censored Times
by Sex
Density
0.006
0.004
0.002
0.000
0
100
200
300
Event time (days)
sex
male
female

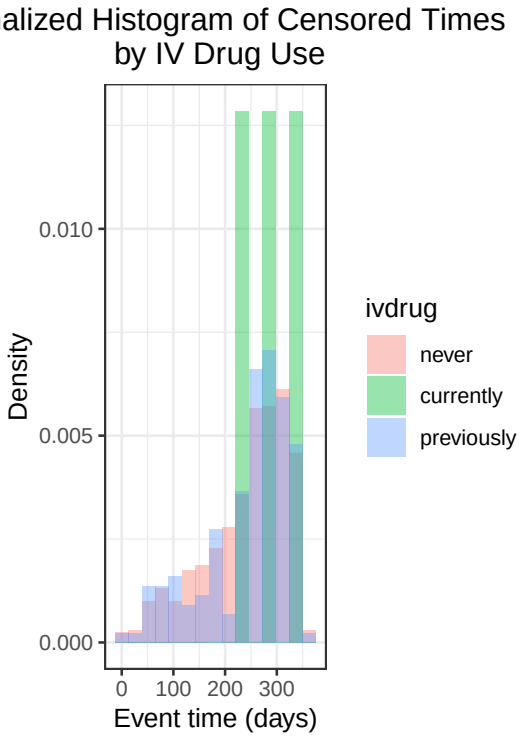

Censored Time Density
by CD4 Stratum
density
0.006
0.004
0.002
0.000
0
100
200
300
time_event
cd4_lvl
<=50
>50

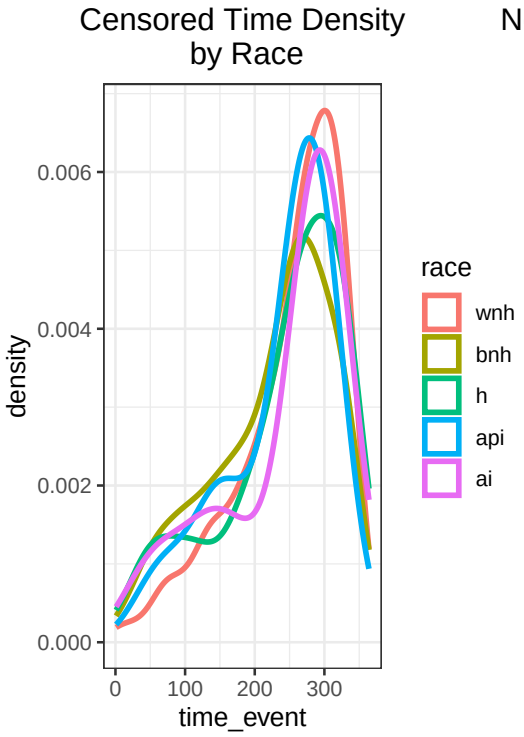

Histogram of Censored Times
by CD4 Stratum
Density
0.006
0.004
0.002
0.000
0
100
200
300
400
Event time (days)
cd4_lvl
<=50
>50

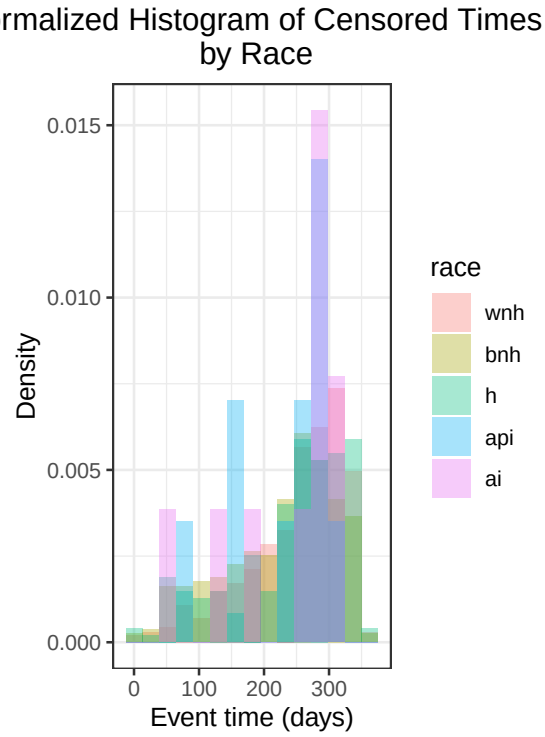

Censored Time Density
by CD4 Stratum
density
0.006
0.004
0.002
0.000
0
100
200
300
time_event
cd4_strata
<=50
51−200
>200

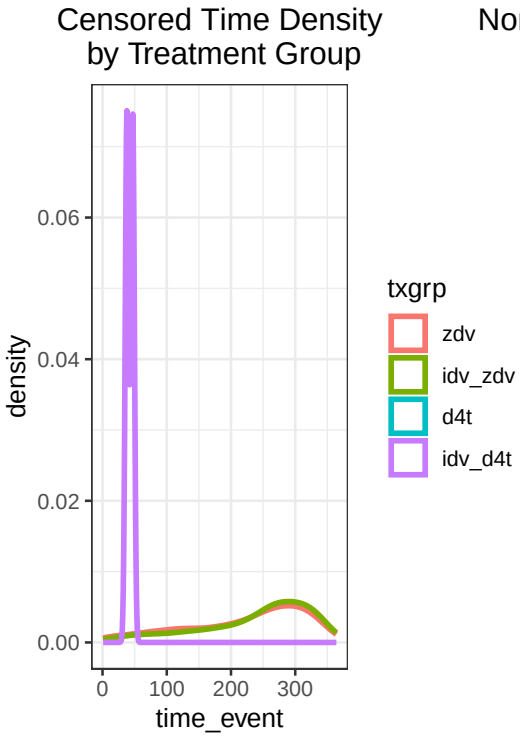

Normalized Histogram of Censored Times
by CD4 Stratum
Density
0.006
0.004
0.002
0.000
0
100
200
300
400
Event time (days)
cd4_lvl
<=50
>50

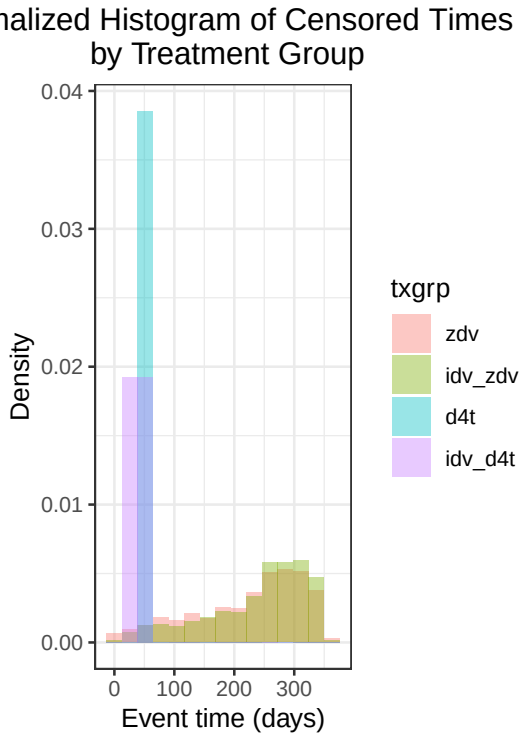

Censored Time Density
by IV Drug Use
density
0.006
0.004
0.002
0.000
0
100
200
300
time_event
ivdrug
never
currently
previously


Normalized Histogram of Censored Times
by IV Drug Use
Density
0.010
0.005
0.000
0
100
200
300
Event time (days)
ivdrug
never
currently
previously










```
library(survival)
library(survminer)
```

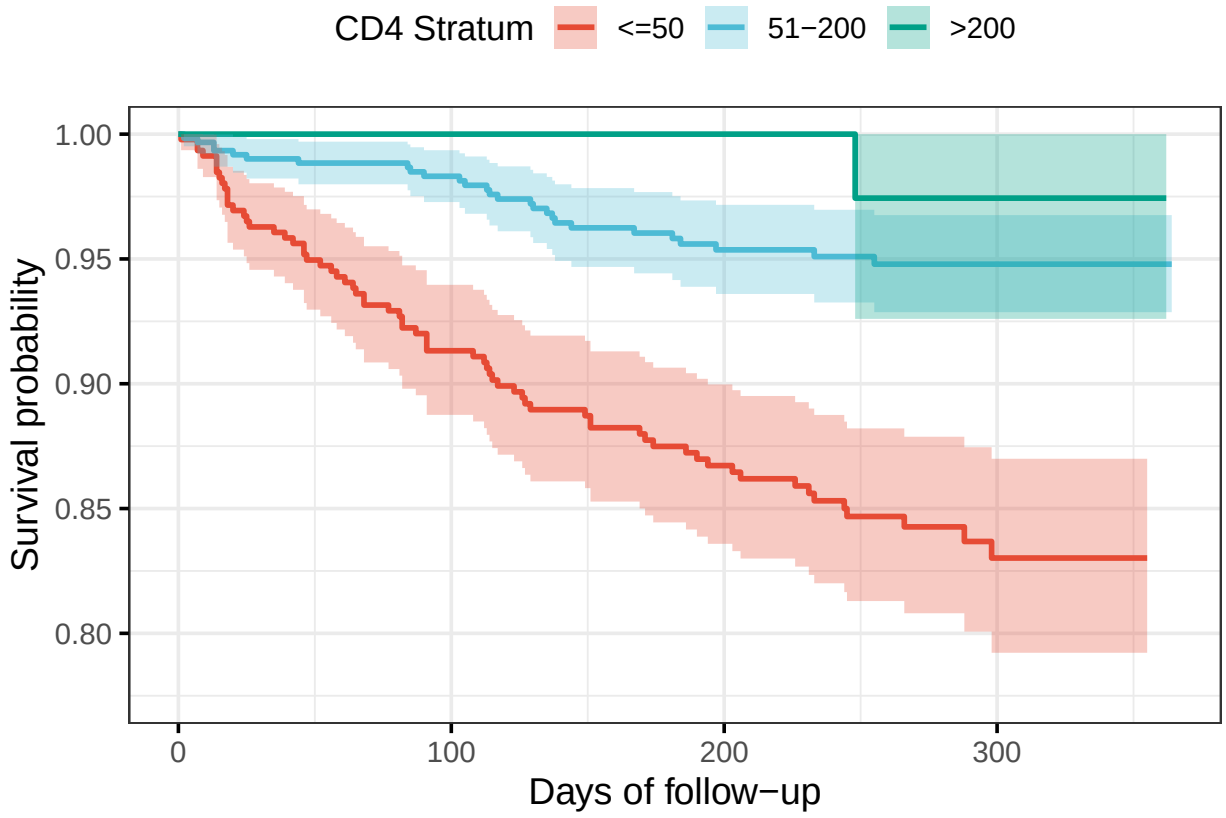
```r
fit <- survfit(Surv(time_event, event) ~ cd4_strata, data = data_new)

ggsurv <- ggsurvplot(
  fit,
  conf.int = TRUE,
  censor = FALSE,                      # Quita las cruces de censura
  risk.table = FALSE,
  palette = c("#E64B35", "#4DBBD5", "#00A087"),
  legend.title = "CD4 Stratum",
  legend.labs = c("<=50", "51-200", ">200"),
  ggtheme = theme_bw(base_size = 14),
  xlab = "Days of follow-up",
  ylab = "Survival probability"
)

# Ajustar eje Y y mejorar el tema
ggsurv$plot <- ggsurv$plot +
  coord_cartesian(ylim = c(0.775, 1)) +     # Recorte para ver mejor diferencias
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    legend.position = "top",
    legend.title = element_text(size = 13),
    legend.text = element_text(size = 11)
  )
```

```
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```

```r
ggsurv
```




```r
library(survival)
library(survminer)

# Modelo Kaplan-Meier para CD4 dicotómico
fit2 <- survfit(Surv(time_event, event) ~ cd4_lvl, data = data_new)

ggsurv2 <- ggsurvplot(
  fit2,
  conf.int = TRUE,
  censor = FALSE,                    # Sin marcas de censura
  risk.table = FALSE,
  palette = c("#E64B35", "#4DBBD5"),  # Solo 2 colores
  legend.title = "CD4 Level",
  legend.labs = c("<=50", ">50"),      # Etiquetas del estrato dicotómico
  ggtheme = theme_bw(base_size = 14),
  xlab = "Days of follow-up",
  ylab = "Survival probability"
)

# Mejoras estéticas y rango del eje Y
ggsurv2$plot <- ggsurv2$plot +
  coord_cartesian(ylim = c(0.775, 1)) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    legend.position = "top",
    legend.title = element_text(size = 13),
```

```
    legend.text = element_text(size = 11)
  )
```

```
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```

```
ggsurv2
```