create_scm

```
library(dplyr)
```

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

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(bnlearn)
```

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

```
library(Rgraphviz)
```

```
## Loading required package: graph
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:bnlearn':
##
##     path, score
```

```
## The following objects are masked from 'package:dplyr':
##
##     combine, intersect, setdiff, union
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##     dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##     grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##     order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##     rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##     union, unique, unsplit, which, which.max, which.min
```

```
##
## Attaching package: 'graph'
```

```
## The following objects are masked from 'package:bnlearn':
##
##     degree, nodes, nodes<-
```

```
## Loading required package: grid
```

```
raw_data <- read.csv('data/compas-scores-two-years.csv')
#head(raw_data)
```

```
# Cleaning used by Propublica https://github.com/propublica/compas-analysis/blob/master/
Compas%20Analysis.ipynb
df <- raw_data %>%
filter(days_b_screening_arrest <= 30) %>%
filter(days_b_screening_arrest >= -30) %>%
filter(is_recid != -1) %>%
filter(c_charge_degree != "O") %>%
filter(score_text != 'N/A') %>%
dplyr::select(sex, age_cat, race
              , juv_fel_count, juv_misd_count, juv_other_count
              , priors_count
              , c_charge_degree #, r_charge_degree
              , two_year_recid) %>%
mutate(total_juv_counts = juv_fel_count + juv_misd_count + juv_other_count) %>%
mutate(prior_adult_counts = priors_count - total_juv_counts) %>%
filter(prior_adult_counts > -1)
#head(df)
```

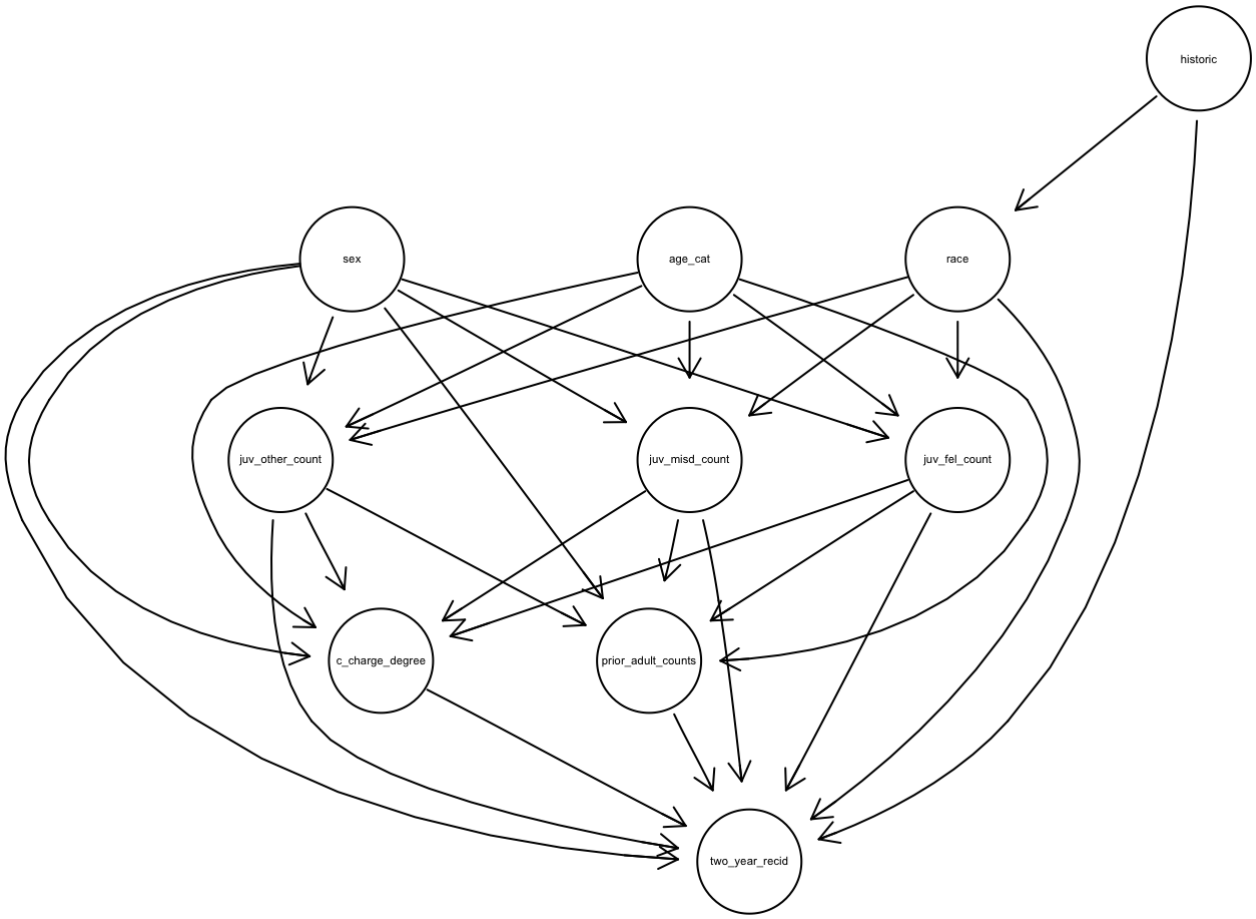
```
df <- df %>%
#  mutate(prior_adult_counts=prior_adult_counts/max(prior_adult_counts)) %>%
#  mutate(juv_fel_count=juv_fel_count/max(total_juv_counts)) %>%
#  mutate(juv_misd_count=juv_misd_count/max(total_juv_counts)) %>%
#  mutate(juv_other_count=juv_other_count/max(total_juv_counts)) %>%
mutate(two_year_recid = factor(two_year_recid)) %>%
select(-c('priors_count','total_juv_counts'))
head(df)
```

```
##    sex         age_cat             race juv_fel_count juv_misd_count
## 1 Male Greater than 45            Other             0              0
## 2 Male         25 - 45 African-American             0              0
## 3 Male    Less than 25 African-American             0              0
## 4 Male         25 - 45            Other             0              0
## 5 Male         25 - 45        Caucasian             0              0
## 6 Male         25 - 45            Other             0              0
##   juv_other_count c_charge_degree two_year_recid prior_adult_counts
## 1               0               F              0                  0
## 2               0               F              1                  0
## 3               1               F              1                  3
## 4               0               M              0                  0
## 5               0               F              1                 14
## 6               0               F              0                  3
```

```
df %>% count(two_year_recid)
```

```
##   two_year_recid    n
## 1              0 3314
## 2              1 2730
```

```
nodes = c("race","sex","historic","age_cat","juv_fel_count","juv_misd_count","juv_other_
count","prior_adult_counts","c_charge_degree","two_year_recid")
e = empty.graph(nodes)
modelstring(e) = "[historic][sex][race|historic][age_cat][juv_fel_count|race:age_cat:se
x][juv_misd_count|race:sex:age_cat][juv_other_count|race:sex:age_cat][prior_adult_counts
|sex:juv_fel_count:juv_misd_count:juv_other_count:age_cat][c_charge_degree|sex:juv_fel_c
ount:juv_misd_count:juv_other_count:age_cat][two_year_recid|race:sex:historic:juv_fel_co
unt:juv_misd_count:juv_other_count:prior_adult_counts:c_charge_degree]"
dag = model2network(modelstring(e), ordering = nodes)
graphviz.plot(dag)
```



All features must be in either decimal point or categorical. All numeric features were converted to categorical.

```
non_historic_dag <- remove.node(dag,"historic") #%>%

df_new <- df %>%
mutate(juv_fel_count, if_else(juv_fel_count > 0, 1, 0)) %>%
mutate(juv_fel_count = factor(juv_fel_count)) %>%

mutate(juv_misd_count, if_else(juv_misd_count > 0, 1, 0)) %>%
mutate(juv_misd_count = factor(juv_misd_count)) %>%

mutate(juv_other_count, if_else(juv_other_count > 0, 1, 0)) %>%
mutate(juv_other_count = factor(juv_other_count)) %>%

mutate(prior_adult_counts = if_else(prior_adult_counts > 0.0, 1, 0)) %>%
mutate(prior_adult_counts = factor(prior_adult_counts)) %>%
select(0:ncol(df))

fitted_dag = bn.fit(x = non_historic_dag, data = df_new)
```

Simplified DAG

```
nodes = c("race","sex","age_cat","juv_fel_count","juv_misd_count","juv_other_count","pri
or_adult_counts","c_charge_degree","two_year_recid")
e = empty.graph(nodes)
modelstring(e) = "[sex][age_cat][juv_fel_count|race:age_cat:sex][juv_misd_count|race:se
x:age_cat][juv_other_count|race:sex:age_cat][prior_adult_counts|juv_fel_count:juv_misd_c
ount:juv_other_count:age_cat][c_charge_degree|juv_fel_count:juv_misd_count:juv_other_cou
nt:age_cat][race][two_year_recid|race:age_cat:prior_adult_counts:c_charge_degree]"

simp_dag = model2network(modelstring(e), ordering = nodes)
graphviz.plot(simp_dag)
```