NIJ_data_cleanup

Anna Evans

2022-06-17

```
#Read in data
nij_full <- read.csv("../BJS/NIJ_s_Recidivism_Challenge_Training_Dataset.csv")
#Missing values converted to NAs
nij_full <- nij_full %>%
  mutate_all(na_if, "")
```

```
#Select relevant variables
nij_df <- nij_full %>% select(Gender, Race, Age_at_Release, Prison_Offense, Prior_Arrest_Episodes_Felony
```

#Recidivism crosstable year 1

```
crosstable(nij_df, c(Race, Gender, Age_at_Release, Prison_Offense, Prior_Arrest_Episodes_Felony), by=Re
  as_flextable()
```

```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

label	variable	Recidivism_Arrest_Year1	
		false	true
Race	BLACK	7115 (68.99%)	3198 (31.01%)
	WHITE	5536 (71.76%)	2179 (28.24%)
Gender	F	1760 (79.39%)	457 (20.61%)
	M	10891 (68.88%)	4920 (31.12%)
Age_at_Release	18-22	844 (58.05%)	610 (41.95%)
	23-27	2306 (63.86%)	1305 (36.14%)
	28-32	2332 (67.61%)	1117 (32.39%)
	33-37	2124 (71.39%)	851 (28.61%)
	38-42	1506 (73.82%)	534 (26.18%)
	43-47	1397 (75.19%)	461 (24.81%)
	48 or older	2142 (81.11%)	499 (18.89%)

label	variable	Recidivism_Arrest_Year1	
		false	true
Prison_Offense	Drug	2681 (74.12%)	936 (25.88%)
	Other	1302 (68.06%)	611 (31.94%)
	Property	3742 (64.55%)	2055 (35.45%)
	Violent/Non-Sex	2789 (73.45%)	1008 (26.55%)
	Violent/Sex	522 (89.54%)	61 (10.46%)
	NA	1615	706
Prior_Arrest_Episodes_Felony	0	144 (83.72%)	28 (16.28%)
	1	1341 (86.68%)	206 (13.32%)
	10 or more	2664 (61.85%)	1643 (38.15%)
	2	1474 (78.57%)	402 (21.43%)
	3	1449 (74.31%)	501 (25.69%)
	4	1306 (70.33%)	551 (29.67%)
	5	1186 (69.85%)	512 (30.15%)
	6	1010 (67.83%)	479 (32.17%)
	7	899 (69.58%)	393 (30.42%)
	8	648 (63.04%)	380 (36.96%)
	9	530 (65.27%)	282 (34.73%)

#Recidivism crosstable year 2

```
crosstable(nij_df, c(Race, Gender, Age_at_Release, Prison_Offense, Prior_Arrest_Episodes_Felony), by=Re
  as_flextable()
```

```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

label	variable	Recidivism_Arrest_Year2	
		false	true
Race	BLACK	8483 (82.26%)	1830 (17.74%)
	WHITE	6292 (81.56%)	1423 (18.44%)
Gender	F	1853 (83.58%)	364 (16.42%)
	M	12922 (81.73%)	2889 (18.27%)
	18-22	1158 (79.64%)	296 (20.36%)

label	variable	Recidivism_Arrest_Year2	
		false	true
Age_at_Release	23-27	2871 (79.51%)	740 (20.49%)
	28-32	2793 (80.98%)	656 (19.02%)
	33-37	2437 (81.92%)	538 (18.08%)
	38-42	1686 (82.65%)	354 (17.35%)
	43-47	1550 (83.42%)	308 (16.58%)
	48 or older	2280 (86.33%)	361 (13.67%)
Prison_Offense	Drug	2969 (82.08%)	648 (17.92%)
	Other	1539 (80.45%)	374 (19.55%)
	Property	4710 (81.25%)	1087 (18.75%)
	Violent/Non-Sex	3148 (82.91%)	649 (17.09%)
	Violent/Sex	505 (86.62%)	78 (13.38%)
	NA	1904	417
Prior_Arrest_Episodes_Felony	0	143 (83.14%)	29 (16.86%)
	1	1354 (87.52%)	193 (12.48%)
	10 or more	3428 (79.59%)	879 (20.41%)
	2	1587 (84.59%)	289 (15.41%)
	3	1624 (83.28%)	326 (16.72%)
	4	1531 (82.44%)	326 (17.56%)
	5	1359 (80.04%)	339 (19.96%)
	6	1192 (80.05%)	297 (19.95%)
	7	1069 (82.74%)	223 (17.26%)
	8	841 (81.81%)	187 (18.19%)
	9	647 (79.68%)	165 (20.32%)

#Recidivism crosstable year 3

```
crosstable(nij_df, c(Race, Gender, Age_at_Release, Prison_Offense, Prior_Arrest_Episodes_Felony), by=Re
  as_flextable()
```

```
## Warning: Warning: fonts used in ‘flextable‘ are ignored because the ‘pdflatex‘
## engine is used and not ‘xelatex‘ or ‘lualatex‘. You can avoid this warning
## by using the ‘set_flextable_defaults(fonts_ignore=TRUE)‘ command or use a
## compatible engine by defining ‘latex_engine: xelatex‘ in the YAML header of the
## R Markdown document.
```

label	variable	Recidivism_Arrest_Year3	
		false	true
Race	BLACK	9265 (89.84%)	1048 (10.16%)
	WHITE	6972 (90.37%)	743 (9.63%)
Gender	F	2027 (91.43%)	190 (8.57%)
	M	14210 (89.87%)	1601 (10.13%)
Age_at_Release	18-22	1314 (90.37%)	140 (9.63%)
	23-27	3249 (89.98%)	362 (10.02%)
	28-32	3085 (89.45%)	364 (10.55%)
	33-37	2654 (89.21%)	321 (10.79%)
	38-42	1831 (89.75%)	209 (10.25%)
	43-47	1692 (91.07%)	166 (8.93%)
	48 or older	2412 (91.33%)	229 (8.67%)
Prison_Offense	Drug	3246 (89.74%)	371 (10.26%)
	Other	1722 (90.02%)	191 (9.98%)
	Property	5228 (90.18%)	569 (9.82%)
	Violent/Non-Sex	3401 (89.57%)	396 (10.43%)
	Violent/Sex	534 (91.60%)	49 (8.40%)
	NA	2106	215
Prior_Arrest_Episodes_Felony	0	155 (90.12%)	17 (9.88%)
	1	1428 (92.31%)	119 (7.69%)
	10 or more	3838 (89.11%)	469 (10.89%)
	2	1707 (90.99%)	169 (9.01%)
	3	1750 (89.74%)	200 (10.26%)
	4	1675 (90.20%)	182 (9.80%)
	5	1546 (91.05%)	152 (8.95%)
	6	1334 (89.59%)	155 (10.41%)
	7	1151 (89.09%)	141 (10.91%)
	8	921 (89.59%)	107 (10.41%)
	9	732 (90.15%)	80 (9.85%)

#Failure rate of all persons

```
#Failure rate for year 1
year1 <- nij_df %>%
  group_by(Recidivism_Arrest_Year1) %>%
  summarize(n = n(), recidivism_arrest_year = 1) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
```

```
  filter(Recidivism_Arrest_Year1 == "true") %>%
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
#Failure rate for year 2
year2 <- nij_df %>%
  group_by(Recidivism_Arrest_Year2) %>%
  filter(Recidivism_Arrest_Year1 == "false") %>%
  summarize(n = n(), recidivism_arrest_year = 2) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year2 == "true") %>%
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
#Failure rate for year 3
year3 <- nij_df %>%
  group_by(Recidivism_Arrest_Year3) %>%
  filter(Recidivism_Arrest_Year2 == "false", Recidivism_Arrest_Year1 == "false") %>%
  summarize(n = n(), recidivism_arrest_year = 3) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year3 == "true") %>%
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
#Combine all failure rates into one data frame
fr_all <- bind_rows(year1, year2, year3) %>%
  select(recidivism_arrest_year, failure_rate, n) %>%
  mutate(category = "all", attribute = NA, .before = recidivism_arrest_year)
```

#Failure rate by category function

```
#Function that creates a data frame of failure rates of all years for each attribute of a given variabl
fr_table <- function(df, x){

x <- enquo(x)
  year1 <- df %>%
  group_by(!!x, Recidivism_Arrest_Year1) %>%
  summarize(n = n(), recidivism_arrest_year = 1) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year1 == "true") %>%
  ungroup()

year2 <- df %>%
  group_by(!!x, Recidivism_Arrest_Year2) %>%
  filter(Recidivism_Arrest_Year1 == "false") %>%
  summarize(n = n(), recidivism_arrest_year = 2) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year2 == "true") %>%
```

```
  ungroup()

year3 <- df %>%
  group_by(!!x, Recidivism_Arrest_Year3) %>%
  filter(Recidivism_Arrest_Year2 == "false", Recidivism_Arrest_Year1 == "false") %>%
  summarize(n = n(), recidivism_arrest_year = 3) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year3 == "true") %>%
  ungroup()

bind_rows(year1, year2, year3) %>%
  select(!!x, recidivism_arrest_year, failure_rate, n) %>%
  mutate(attribute = !!x, .before = recidivism_arrest_year) %>%
  select(attribute, recidivism_arrest_year, failure_rate, n) -> fr_df

return(fr_df)
}
```

#Failure rate dataframe

```
#Combining all failure rates of each variable
fr_table(nij_df, Gender) %>% mutate(category = "Gender", .before = attribute) -> fr_gender
```

```
## `summarise()` regrouping output by 'Gender' (override with `.groups` argument)
## `summarise()` regrouping output by 'Gender' (override with `.groups` argument)
## `summarise()` regrouping output by 'Gender' (override with `.groups` argument)
```

```
fr_table(nij_df, Race) %>% mutate(category = "Race", .before = attribute)->fr_race
```

```
## `summarise()` regrouping output by 'Race' (override with `.groups` argument)

## `summarise()` regrouping output by 'Race' (override with `.groups` argument)
## `summarise()` regrouping output by 'Race' (override with `.groups` argument)
```

```
fr_table(nij_df, Age_at_Release) %>% mutate(category = "Age_at_Release", .before = attribute) ->fr_age
```

```
## `summarise()` regrouping output by 'Age_at_Release' (override with `.groups` argument)

## `summarise()` regrouping output by 'Age_at_Release' (override with `.groups` argument)
## `summarise()` regrouping output by 'Age_at_Release' (override with `.groups` argument)
```

```
fr_table(nij_df, Prison_Offense) %>% mutate(category = "Prison_Offense", .before = attribute) ->fr_offe
```

```
## `summarise()` regrouping output by 'Prison_Offense' (override with `.groups` argument)

## `summarise()` regrouping output by 'Prison_Offense' (override with `.groups` argument)
## `summarise()` regrouping output by 'Prison_Offense' (override with `.groups` argument)
```

```
fr_table(nij_df, Prior_Arrest_Episodes_Felony) %>% mutate(category = "Prior_Arrest_Episodes_Felony", .b
```

```
## `summarise()` regrouping output by 'Prior_Arrest_Episodes_Felony' (override with `.groups` argument)

## `summarise()` regrouping output by 'Prior_Arrest_Episodes_Felony' (override with `.groups` argument)
## `summarise()` regrouping output by 'Prior_Arrest_Episodes_Felony' (override with `.groups` argument)
```

```
bind_rows(fr_all, fr_gender, fr_race, fr_age, fr_offense, fr_priorar) -> fr_by_cat
```

#Cumulative recidivism for all persons

```
 year1 <- nij_df %>%
  group_by(Recidivism_Arrest_Year1) %>%
  summarize(n = n(), recidivism_arrest_year = 1) %>%
  mutate(cum_recid = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year1 == "true") %>%
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
year2 <- nij_df %>%
  group_by(Recidivism_Arrest_Year2) %>%
  summarize(n = n(), recidivism_arrest_year = 2) %>%
  mutate(failure_rate_yr2 = n/sum(n)*100 ) %>%
  filter(Recidivism_Arrest_Year2 == "true") %>%
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
year2['year1fr'] <- year1$cum_recid
year2 %>% mutate(cum_recid = failure_rate_yr2 + year1fr) -> year2

year3 <- nij_df %>%
  group_by(Recidivism_Arrest_Year3) %>%
  summarize(n = n(), recidivism_arrest_year = 3) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year3 == "true") %>%
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
year3['year2cum'] <- year2$cum_recid
year3 %>% mutate(cum_recid = failure_rate + year2cum) -> year3

bind_rows(year1, year2, year3) %>%
  select(recidivism_arrest_year, cum_recid, n) %>%
  mutate(attribute = "all") -> cum_recid
```

#Cumulative recidivism by category

```r
#Function that creates a dataframe of the cumulative recidivism rate for all years and attributes of a
cum_recid_table <- function(df, x){
  x <- enquo(x)
  year1 <- df %>%
  group_by(!!x, Recidivism_Arrest_Year1) %>%
  summarize(n = n(), recidivism_arrest_year = 1) %>%
  mutate(cum_recid = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year1 == "true") %>%
  ungroup()

year2 <- df %>%
  group_by(!!x, Recidivism_Arrest_Year2) %>%
  summarize(n = n(), recidivism_arrest_year = 2) %>%
  mutate(failure_rate_yr2 = n/sum(n)*100 ) %>%
  filter(Recidivism_Arrest_Year2 == "true") %>%
  ungroup()

year2['year1fr'] <- year1$cum_recid
year2 %>% mutate(cum_recid = failure_rate_yr2 + year1fr) -> year2

year3 <- df %>%
  group_by(!!x, Recidivism_Arrest_Year3) %>%
  summarize(n = n(), recidivism_arrest_year = 3) %>%
  mutate(failure_rate = n/sum(n)*100) %>%
  filter(Recidivism_Arrest_Year3 == "true") %>%
  ungroup()

year3['year2cum'] <- year2$cum_recid
year3 %>% mutate(cum_recid = failure_rate + year2cum) -> year3

bind_rows(year1, year2, year3) %>%
  select(!!x, recidivism_arrest_year, cum_recid, n) %>%
  rename(attribute = !!x) -> cr_df
return(cr_df)
}
#Combining all cumulative recidivism rates by variable
cum_recid_table(nij_df, Gender) -> cr_gender
```

```
## `summarise()` regrouping output by 'Gender' (override with `.groups` argument)
## `summarise()` regrouping output by 'Gender' (override with `.groups` argument)
## `summarise()` regrouping output by 'Gender' (override with `.groups` argument)
```

```r
cum_recid_table(nij_df, Race) -> cr_race
```

```
## `summarise()` regrouping output by 'Race' (override with `.groups` argument)

## `summarise()` regrouping output by 'Race' (override with `.groups` argument)
## `summarise()` regrouping output by 'Race' (override with `.groups` argument)
```

```r
cum_recid_table(nij_df, Age_at_Release) -> cr_age
```

```
## `summarise()` regrouping output by 'Age_at_Release' (override with `.groups` argument)
```

```
## `summarise()` regrouping output by 'Age_at_Release' (override with `.groups` argument)
## `summarise()` regrouping output by 'Age_at_Release' (override with `.groups` argument)
```

```
cum_recid_table(nij_df, Prison_Offense) -> cr_offense
```

```
## `summarise()` regrouping output by 'Prison_Offense' (override with `.groups` argument)

## `summarise()` regrouping output by 'Prison_Offense' (override with `.groups` argument)
## `summarise()` regrouping output by 'Prison_Offense' (override with `.groups` argument)
```

```
cum_recid_table(nij_df, Prior_Arrest_Episodes_Felony) -> cr_priorar
```

```
## `summarise()` regrouping output by 'Prior_Arrest_Episodes_Felony' (override with `.groups` argument)

## `summarise()` regrouping output by 'Prior_Arrest_Episodes_Felony' (override with `.groups` argument)
## `summarise()` regrouping output by 'Prior_Arrest_Episodes_Felony' (override with `.groups` argument)
```

```
bind_rows(cum_recid, cr_gender, cr_race, cr_age, cr_offense, cr_priorar) -> cr_by_cat
```

#Final df and create csv

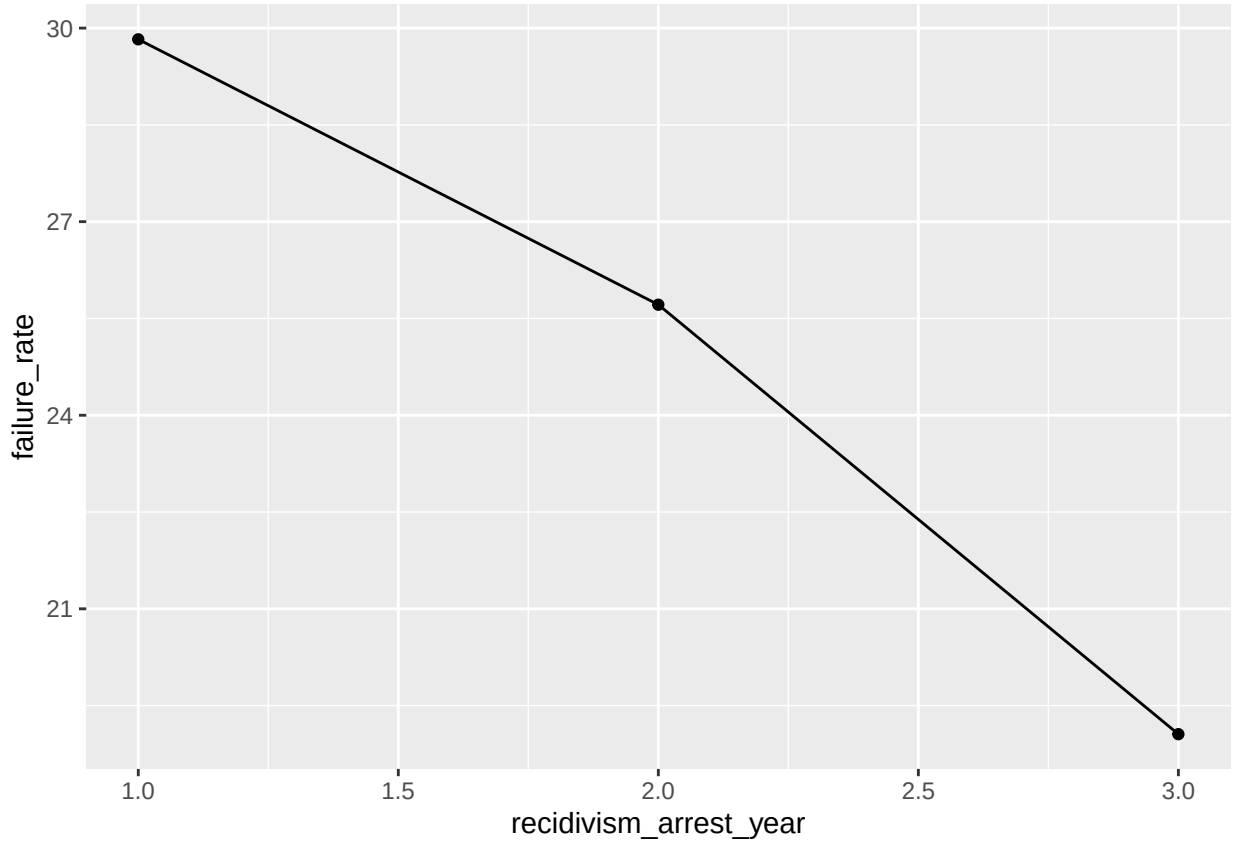
```
#Joining cumulative recidivism rates and annual failure rates into one table
left_join(fr_by_cat, cr_by_cat, by = c("attribute", "recidivism_arrest_year")) -> nij_fr_cr
nij_fr_cr
```

```
## # A tibble: 87 x 7
##    category attribute recidivism_arrest_year failure_rate   n.x cum_recid   n.y
##    <chr>    <chr>                     <dbl>        <dbl> <int>     <dbl> <int>
##  1 all      <NA>                          1         29.8  5377      30.4   706
##  2 all      <NA>                          2         25.7  3253      48.4   417
##  3 all      <NA>                          3         19.1  1791      57.6   215
##  4 Gender   F                             1         20.6   457      20.6   457
##  5 Gender   M                             1         31.1  4920      31.1  4920
##  6 Gender   F                             2         20.7   364      37.0   364
##  7 Gender   M                             2         26.5  2889      49.4  2889
##  8 Gender   F                             3         13.6   190      45.6   190
##  9 Gender   M                             3         20.0  1601      59.5  1601
## 10 Race     BLACK                         1         31.0  3198      31.0  3198
## # ... with 77 more rows
```

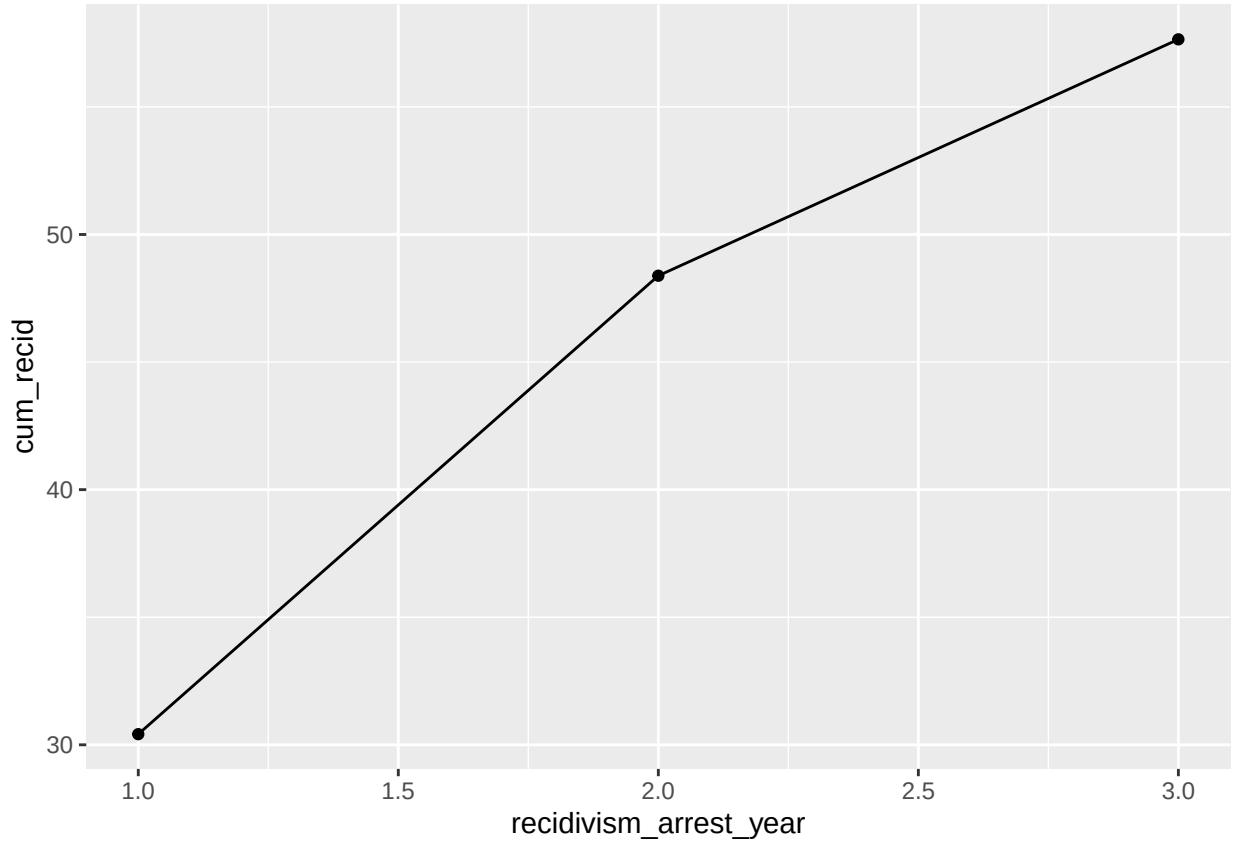
```
write.csv(nij_fr_cr, "nij_fr_cr.csv")
```

#All

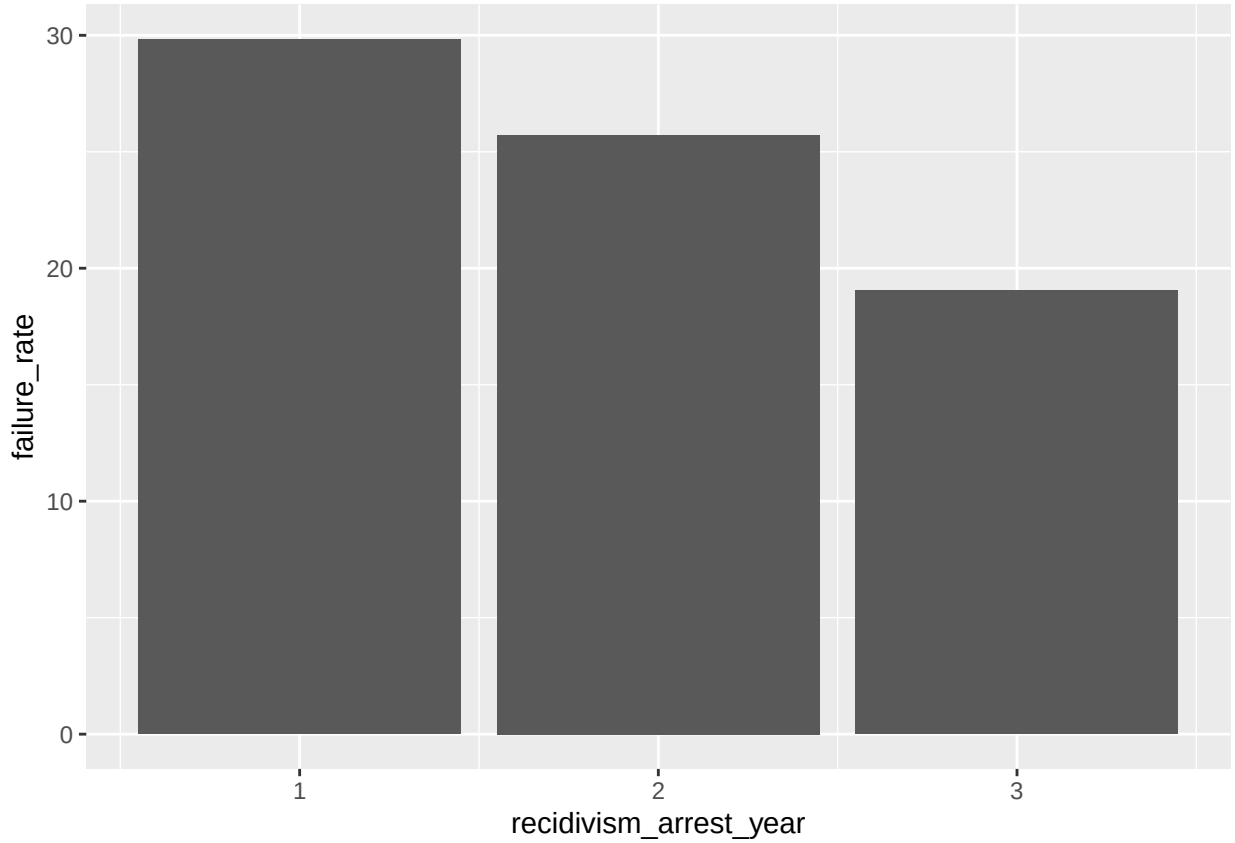
```
nij_fr_cr %>%
  filter(category == "all") %>%
  ggplot(aes(recidivism_arrest_year, failure_rate)) +
  geom_line() +
  geom_point()
```



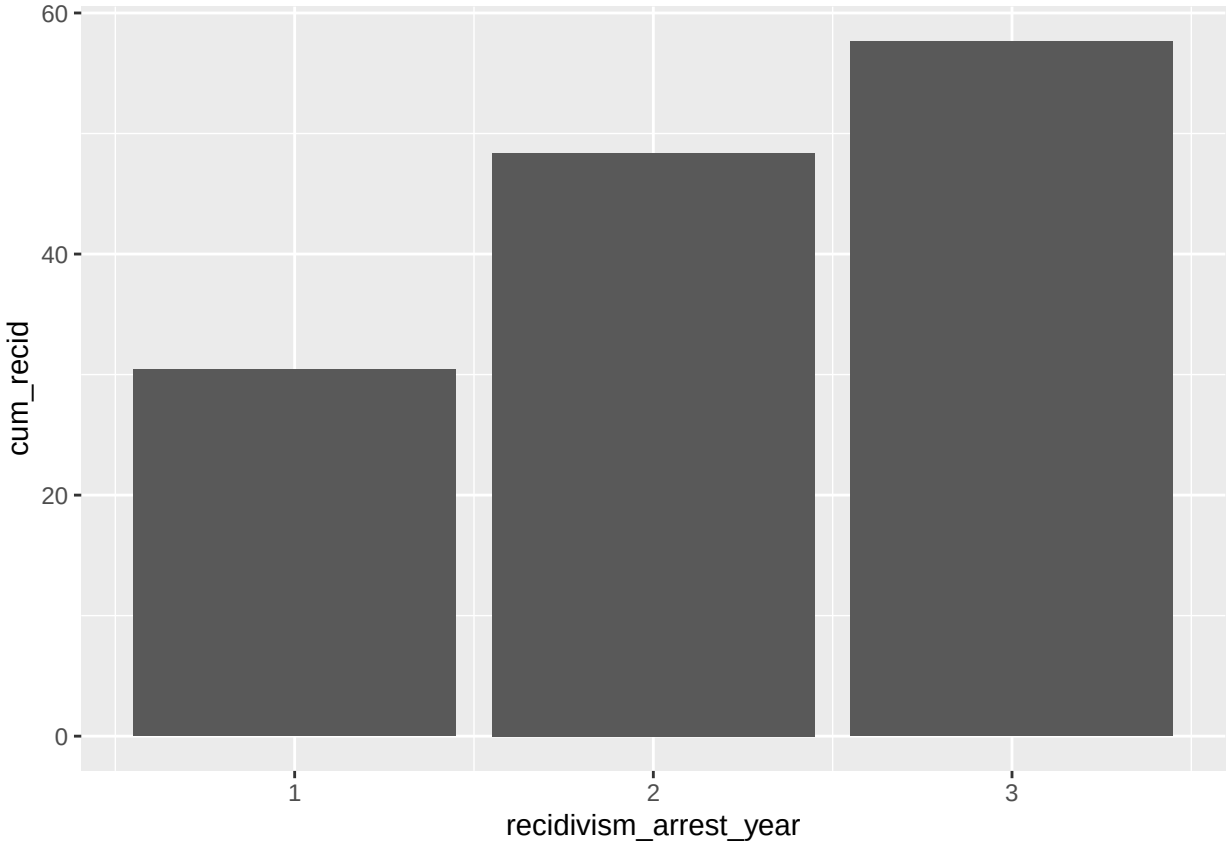

```
nij_fr_cr %>%
  filter(category == "all") %>%
  ggplot(aes(recidivism_arrest_year, cum_recid)) +
  geom_line() +
  geom_point()
```




```
#Bar graph
nij_fr_cr %>%
  filter(category == "all") %>%
  ggplot() +
  geom_bar(aes(x = recidivism_arrest_year, y = failure_rate), stat="identity")
```

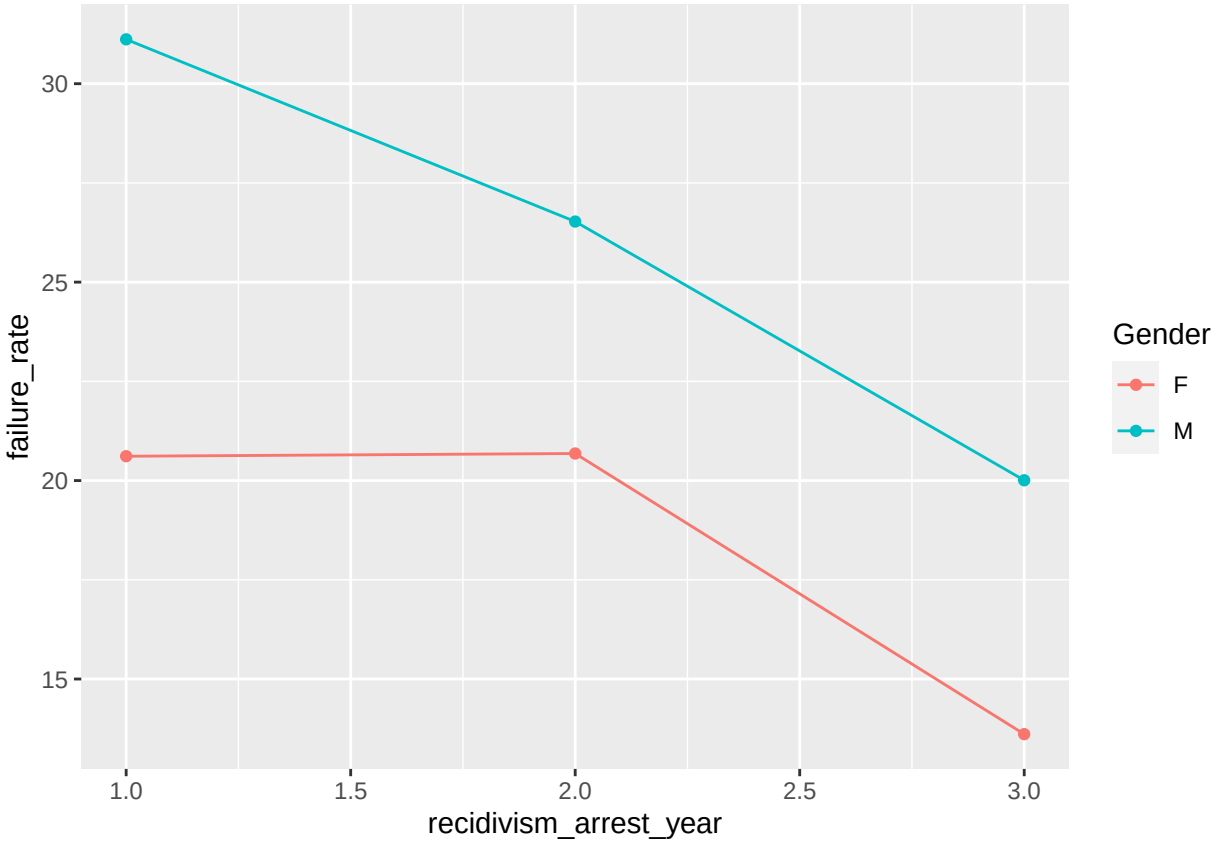



```
nij_fr_cr %>%
  filter(category == "all") %>%
  ggplot() +
  geom_bar(aes(x = recidivism_arrest_year, y = cum_recid), stat="identity")
```

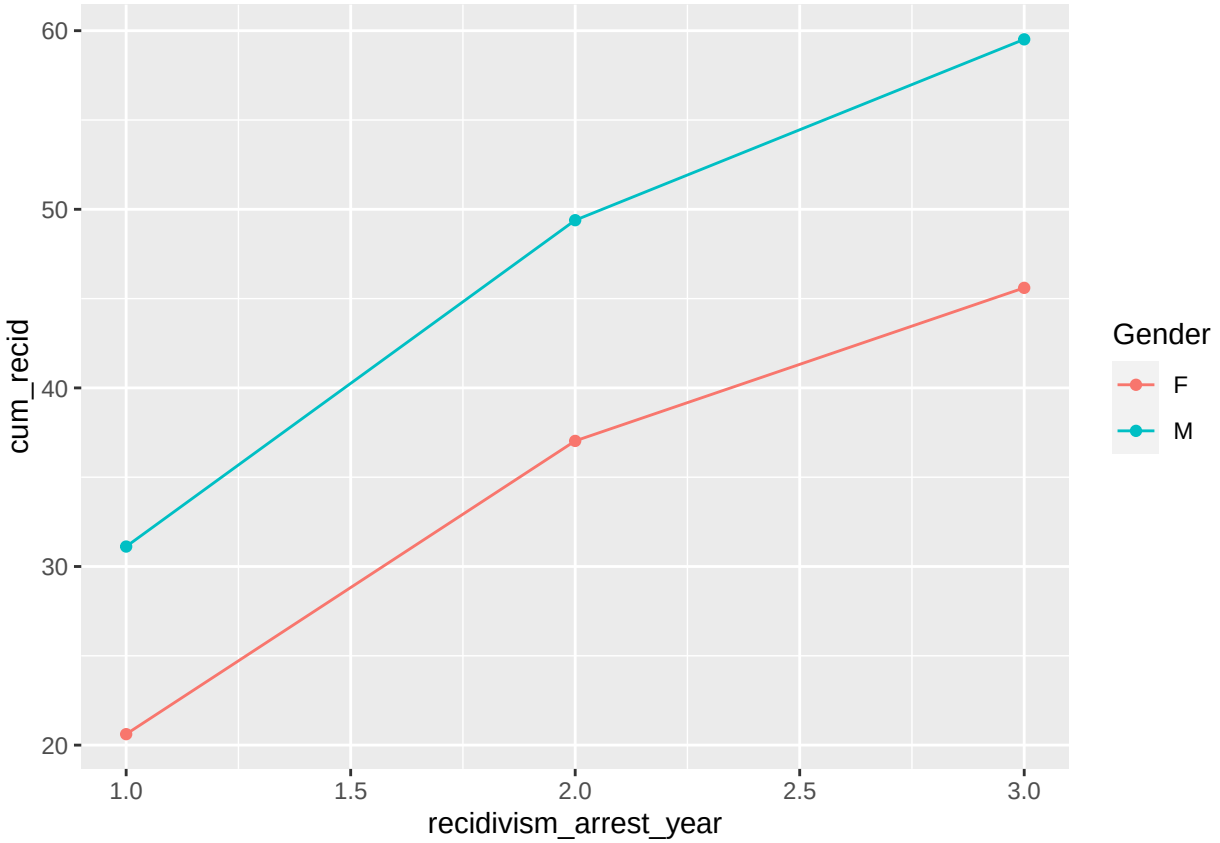


#Gender

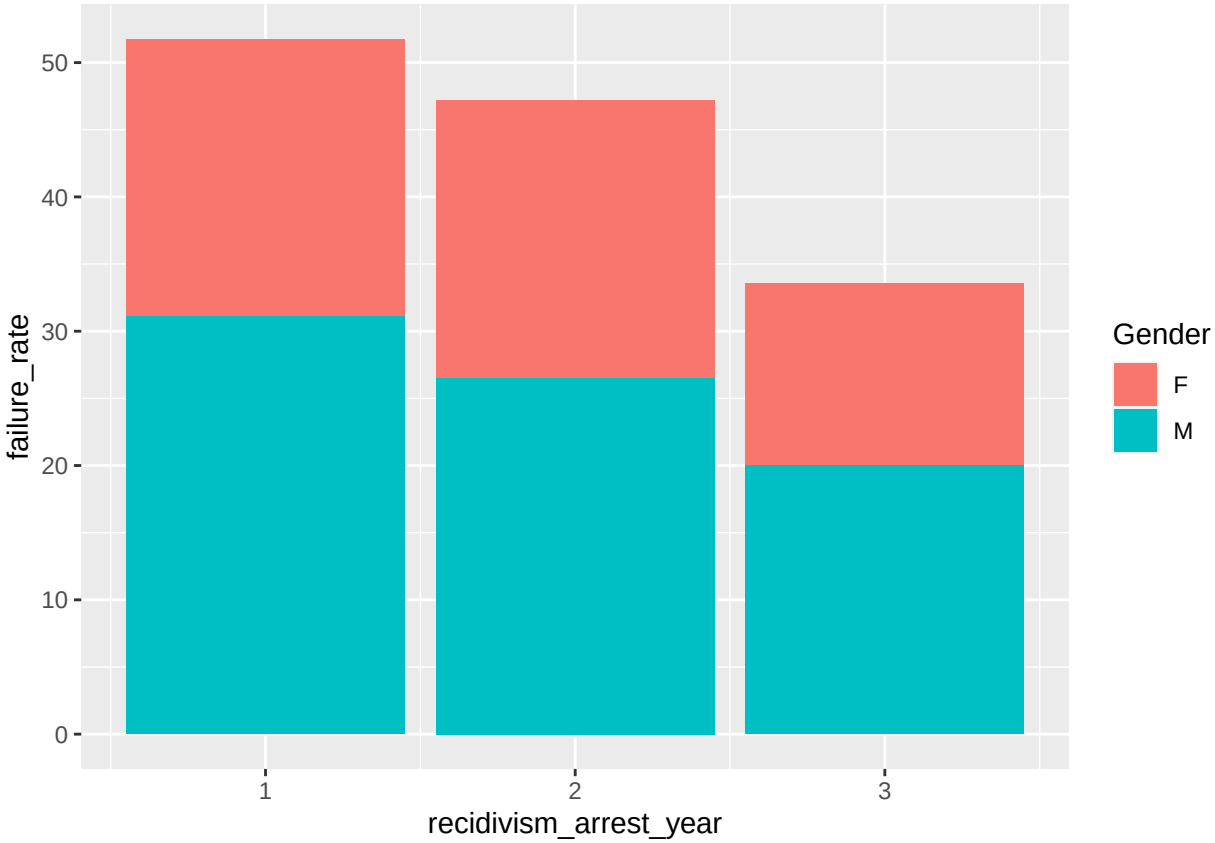
```
nij_fr_cr %>%
  filter(category == "Gender") %>%
  ggplot(aes(recidivism_arrest_year, failure_rate, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Gender")+
  geom_point()
```



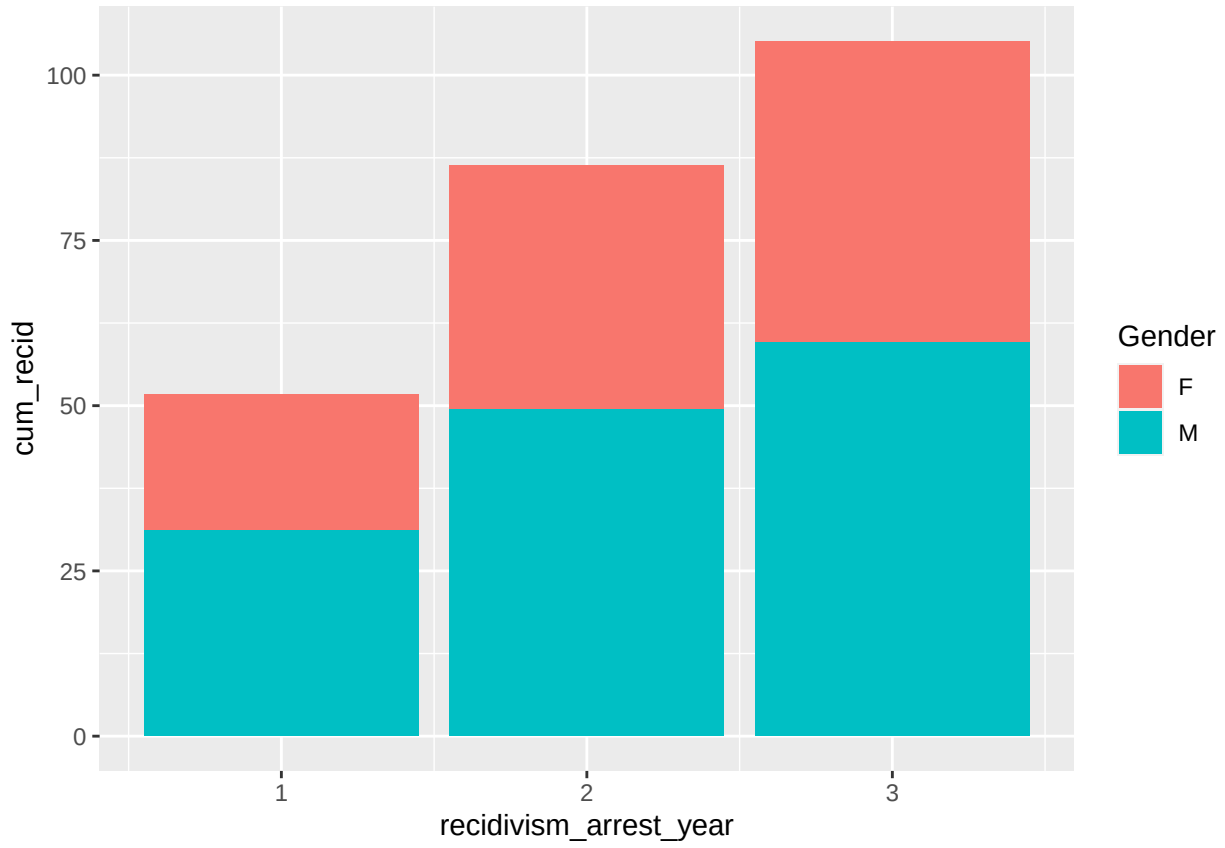

```
nij_fr_cr %>%
  filter(category == "Gender") %>%
  ggplot(aes(recidivism_arrest_year, cum_recid, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Gender")+
  geom_point()
```




```
#Stacked bar graph
nij_fr_cr %>%
  filter(category == "Gender") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = failure_rate), position="stack", stat=
  scale_fill_discrete(name="Gender")
```

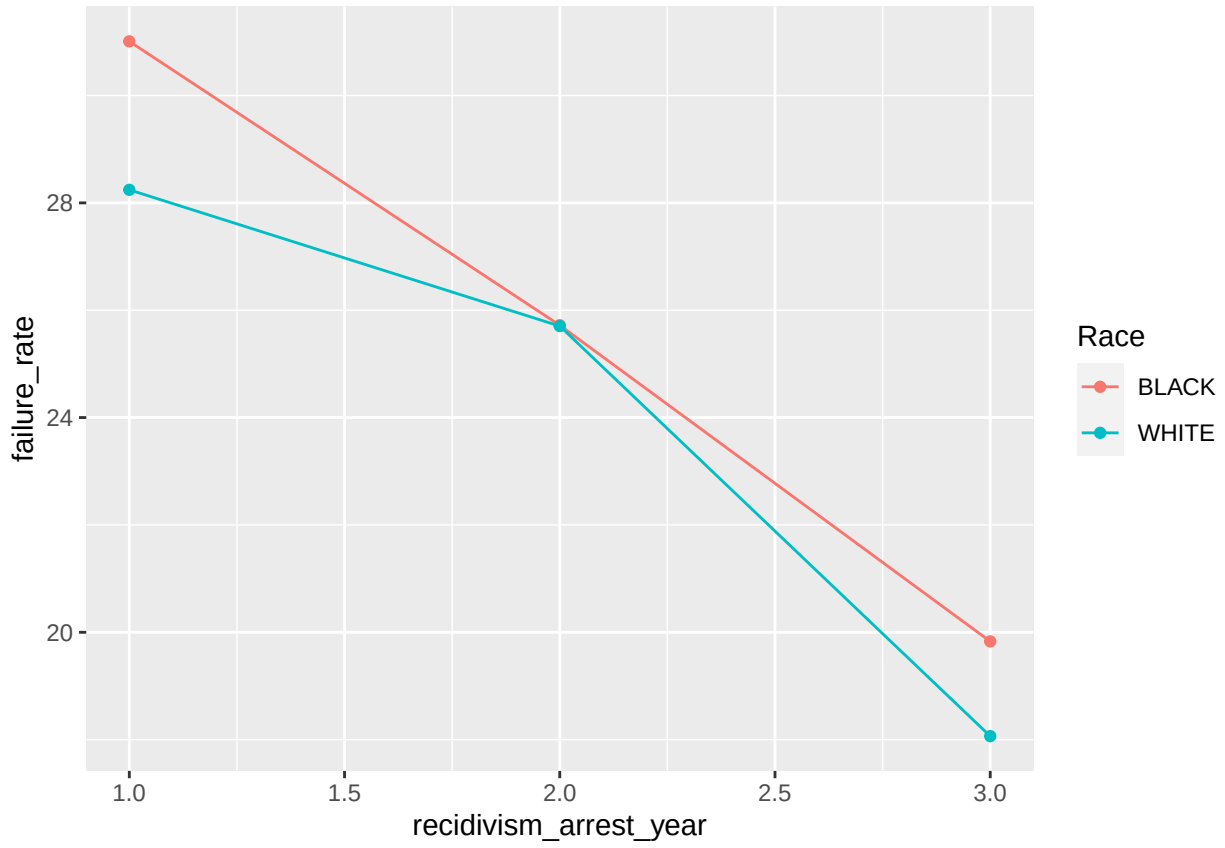



```
nij_fr_cr %>%
  filter(category == "Gender") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = cum_recid), position="stack", stat="id
  scale_fill_discrete(name="Gender")
```

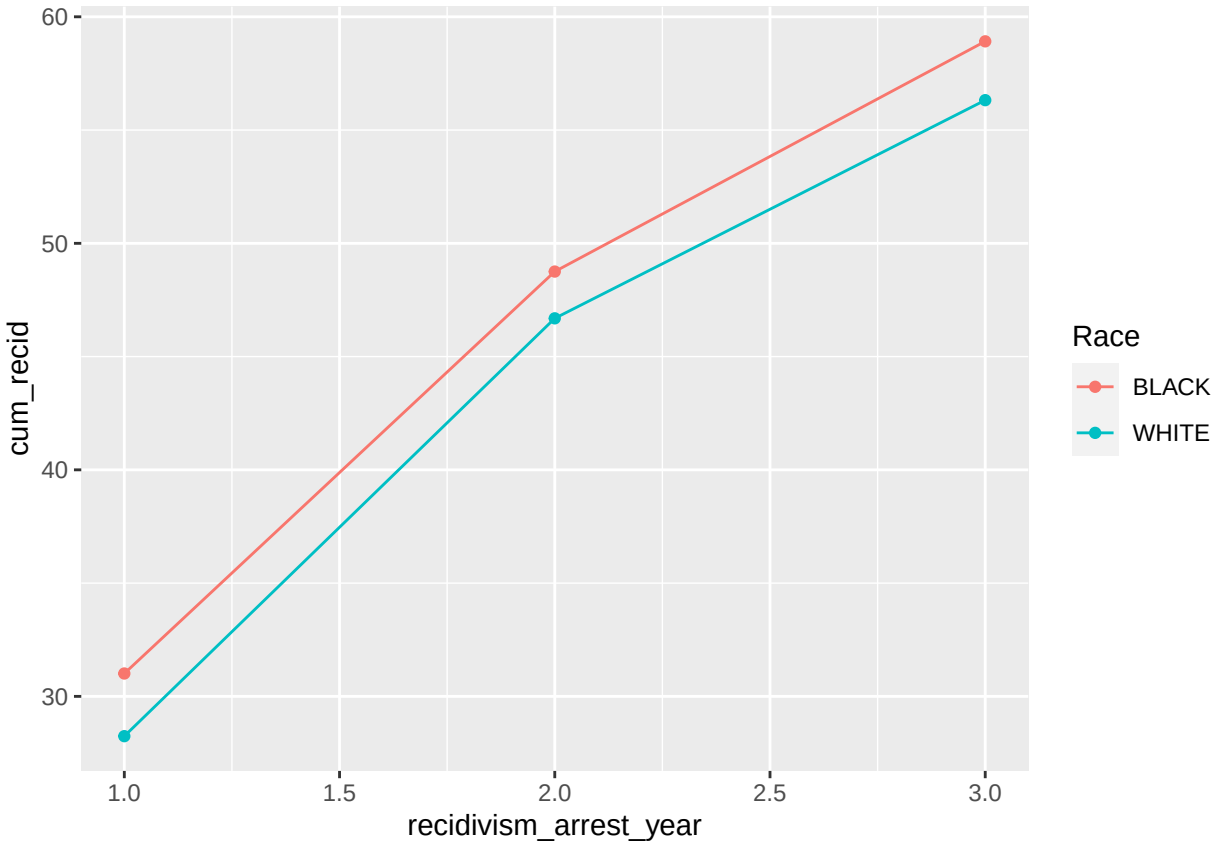


#Race

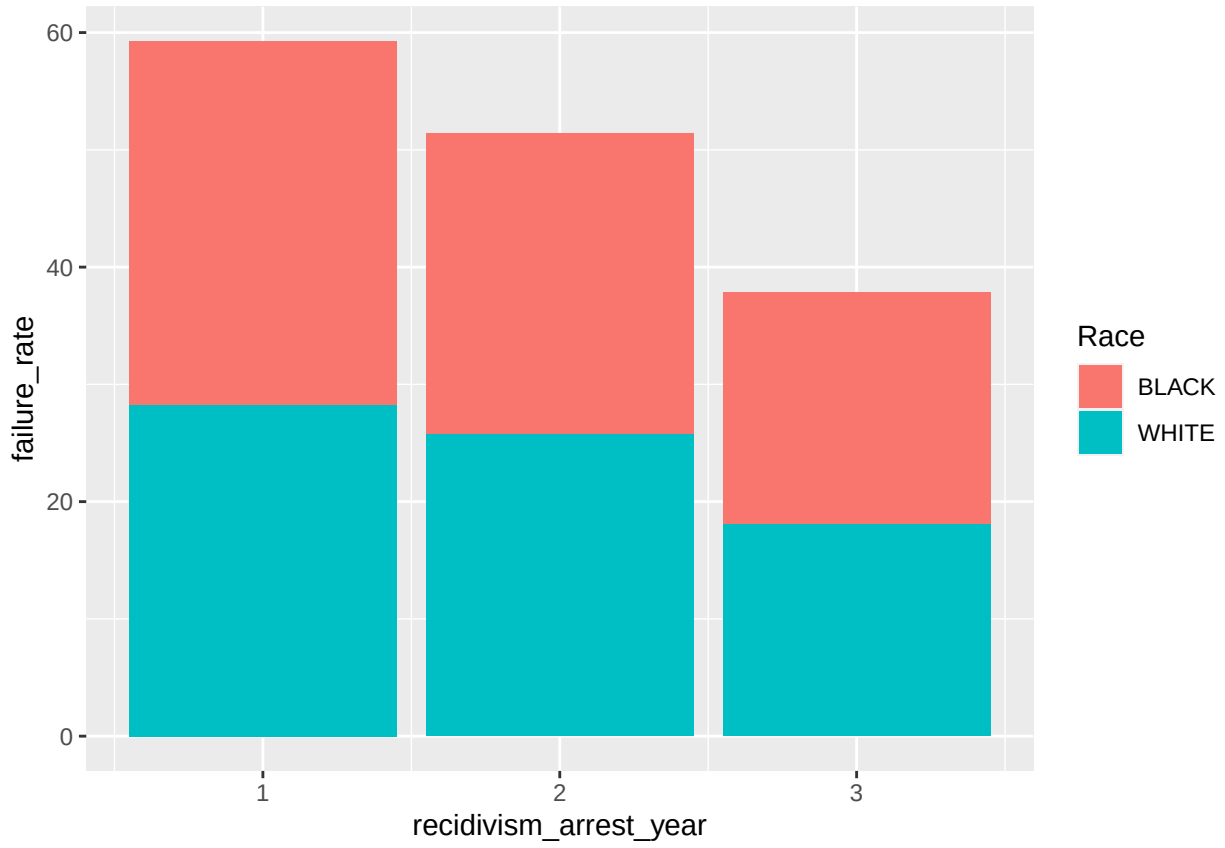
```
#Line graphs
nij_fr_cr %>%
  filter(category == "Race") %>%
  ggplot(aes(recidivism_arrest_year, failure_rate, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Race")+
  geom_point()
```



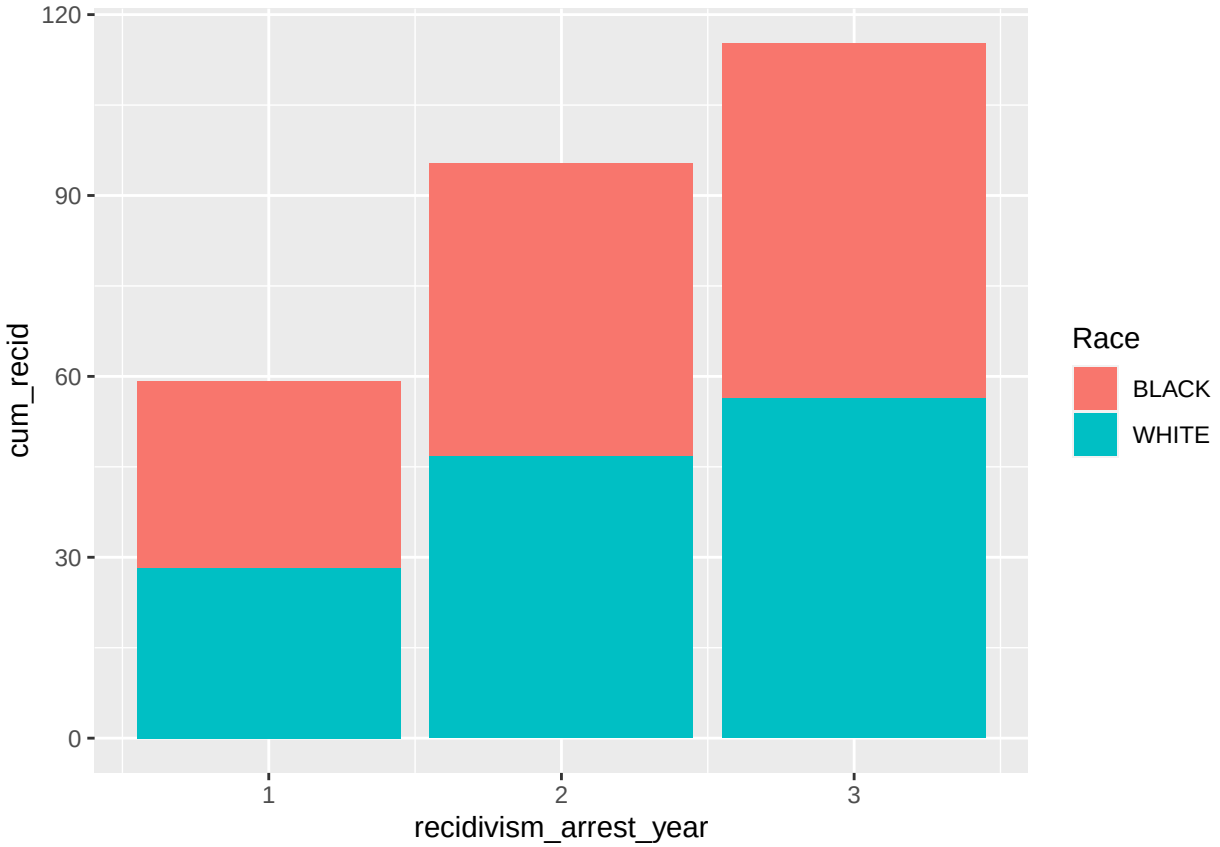

```
nij_fr_cr %>%
  filter(category == "Race") %>%
  ggplot(aes(recidivism_arrest_year, cum_recid, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Race")+
  geom_point()
```




```
#Stacked bar graph
nij_fr_cr %>%
  filter(category == "Race") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = failure_rate), position="stack", stat=
  scale_fill_discrete(name="Race")
```

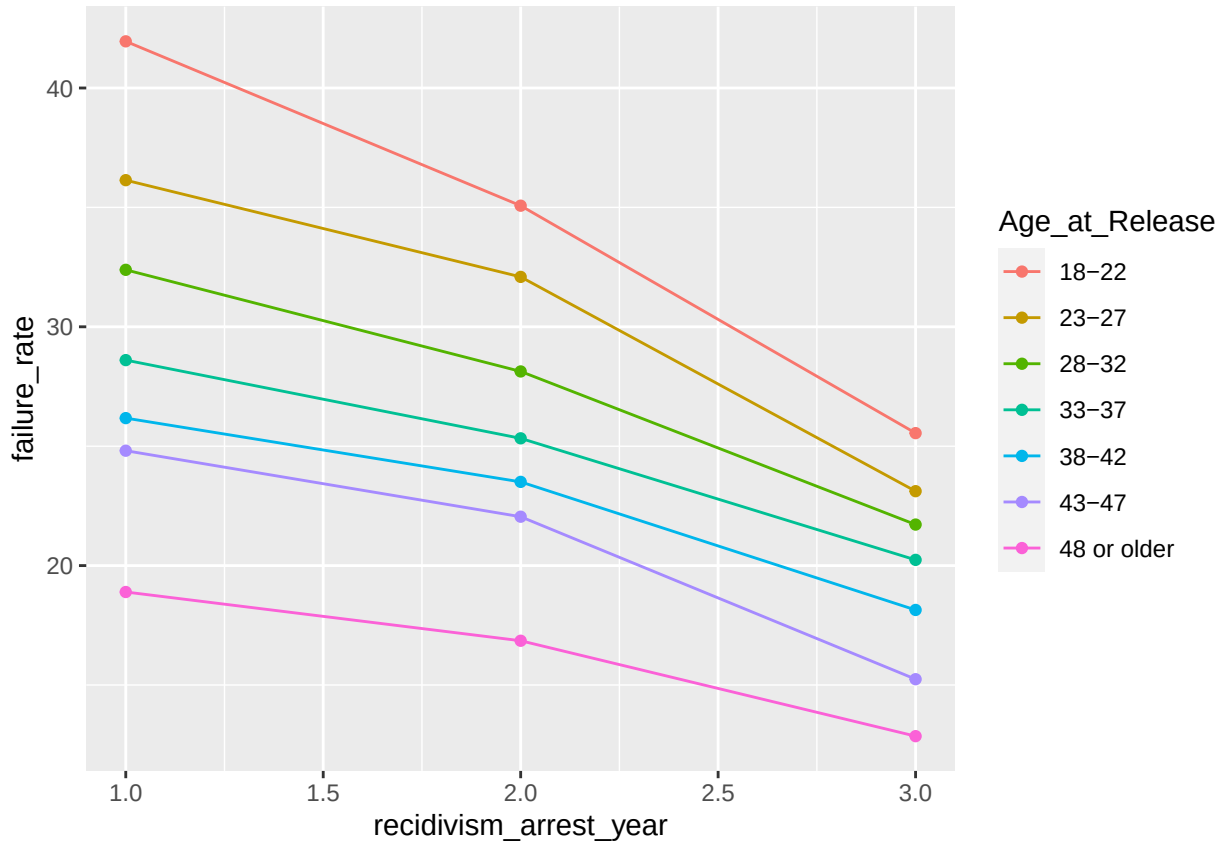



```
nij_fr_cr %>%
  filter(category == "Race") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = cum_recid), position="stack", stat="id
  scale_fill_discrete(name="Race")
```

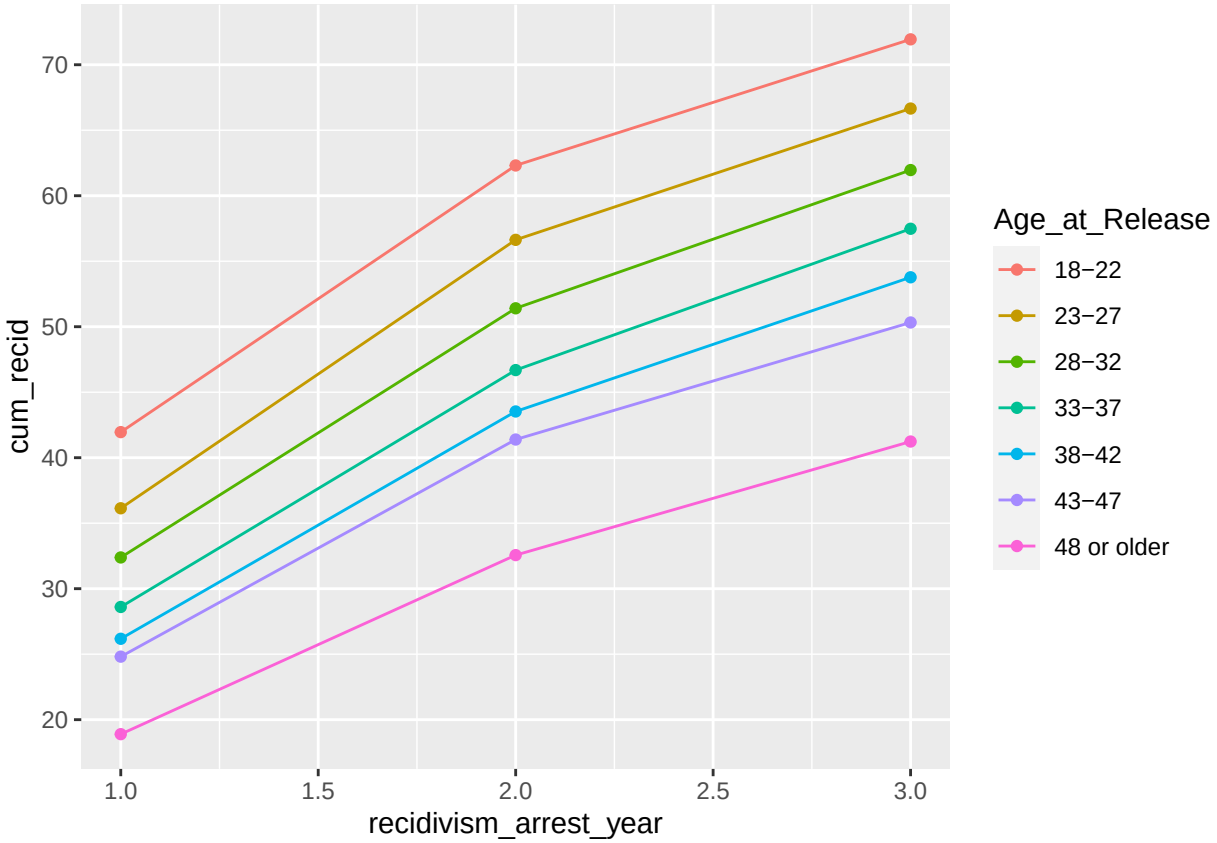


#Age at release

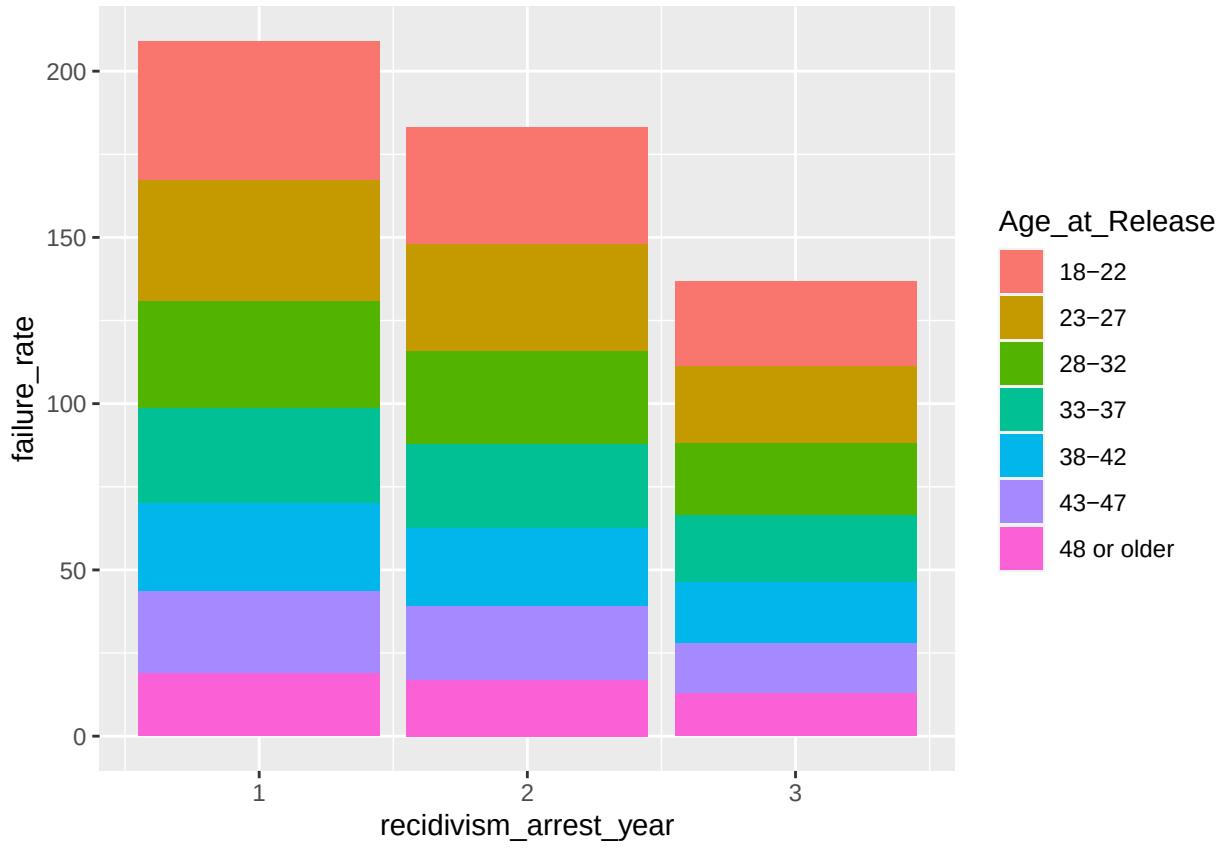
```
nij_fr_cr %>%
  filter(category == "Age_at_Release") %>%
  ggplot(aes(recidivism_arrest_year, failure_rate, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Age_at_Release")+
  geom_point()
```



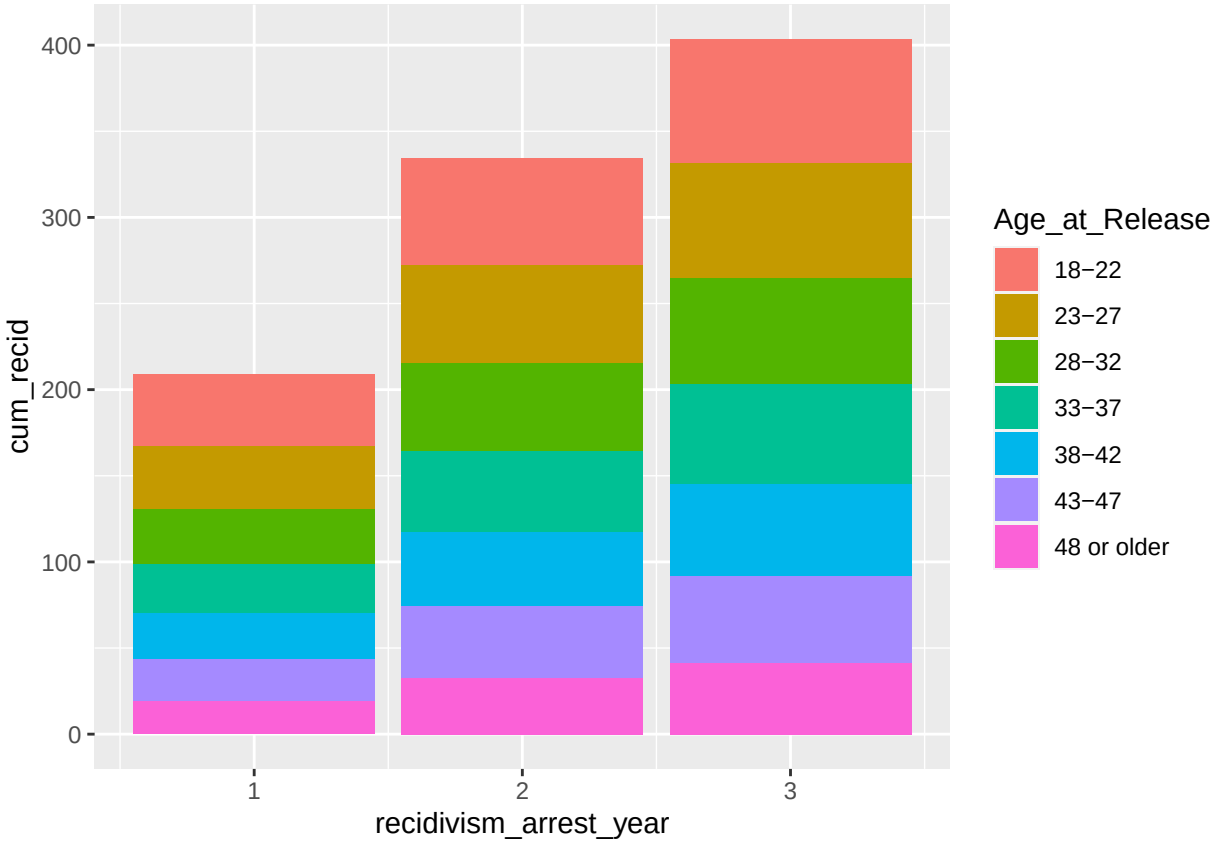

```
nij_fr_cr %>%
  filter(category == "Age_at_Release") %>%
  ggplot(aes(recidivism_arrest_year, cum_recid, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Age_at_Release")+
  geom_point()
```




```
#Stacked bar graph
nij_fr_cr %>%
  filter(category == "Age_at_Release") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = failure_rate), position="stack", stat=
  scale_fill_discrete(name="Age_at_Release")
```

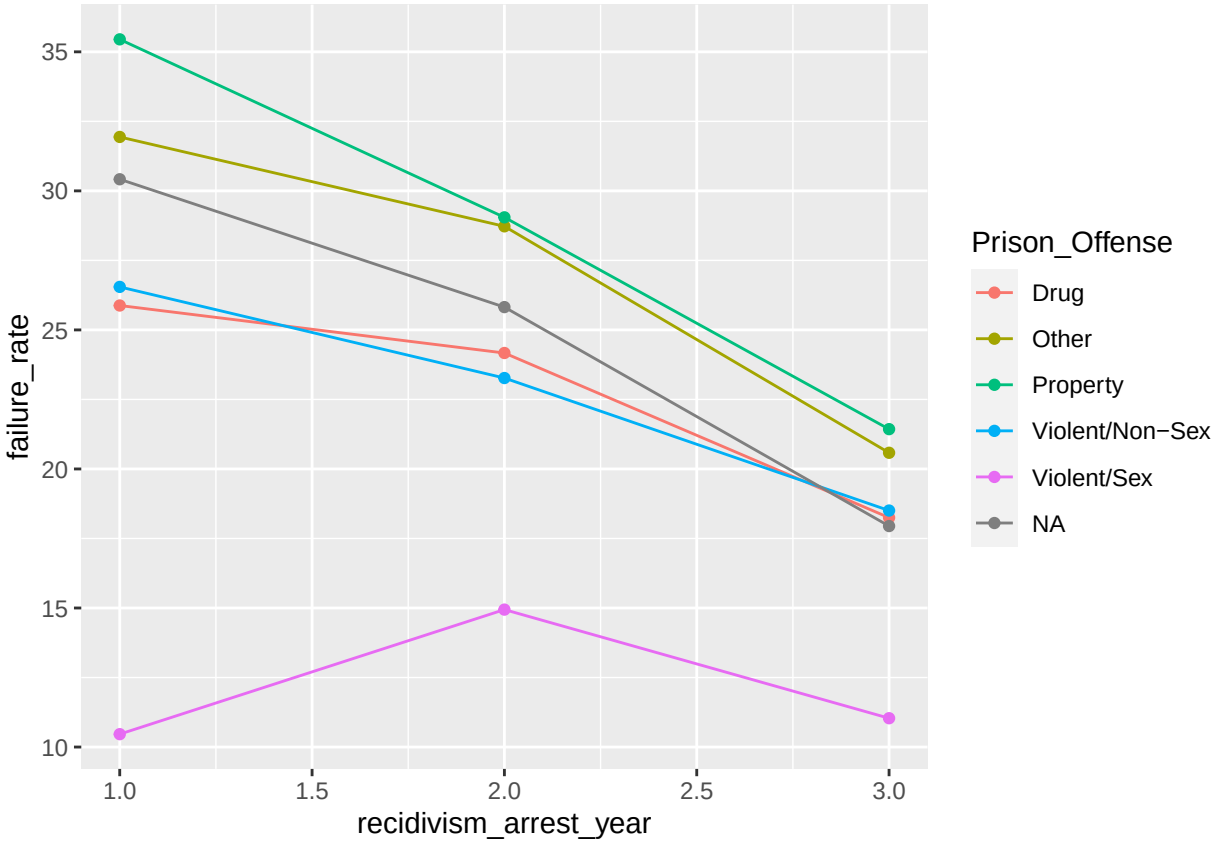



```
nij_fr_cr %>%
  filter(category == "Age_at_Release") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = cum_recid), position="stack", stat="id
  scale_fill_discrete(name="Age_at_Release")
```

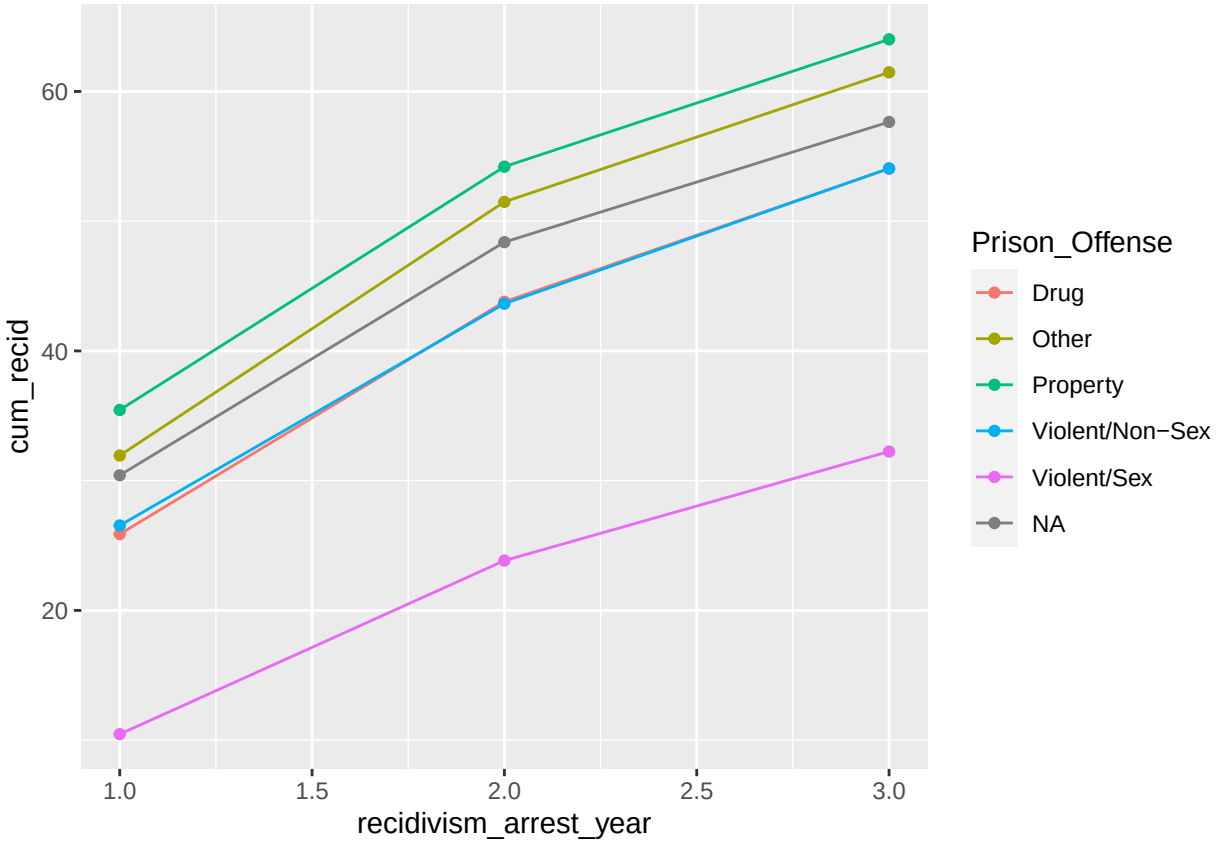


#Prison Offense

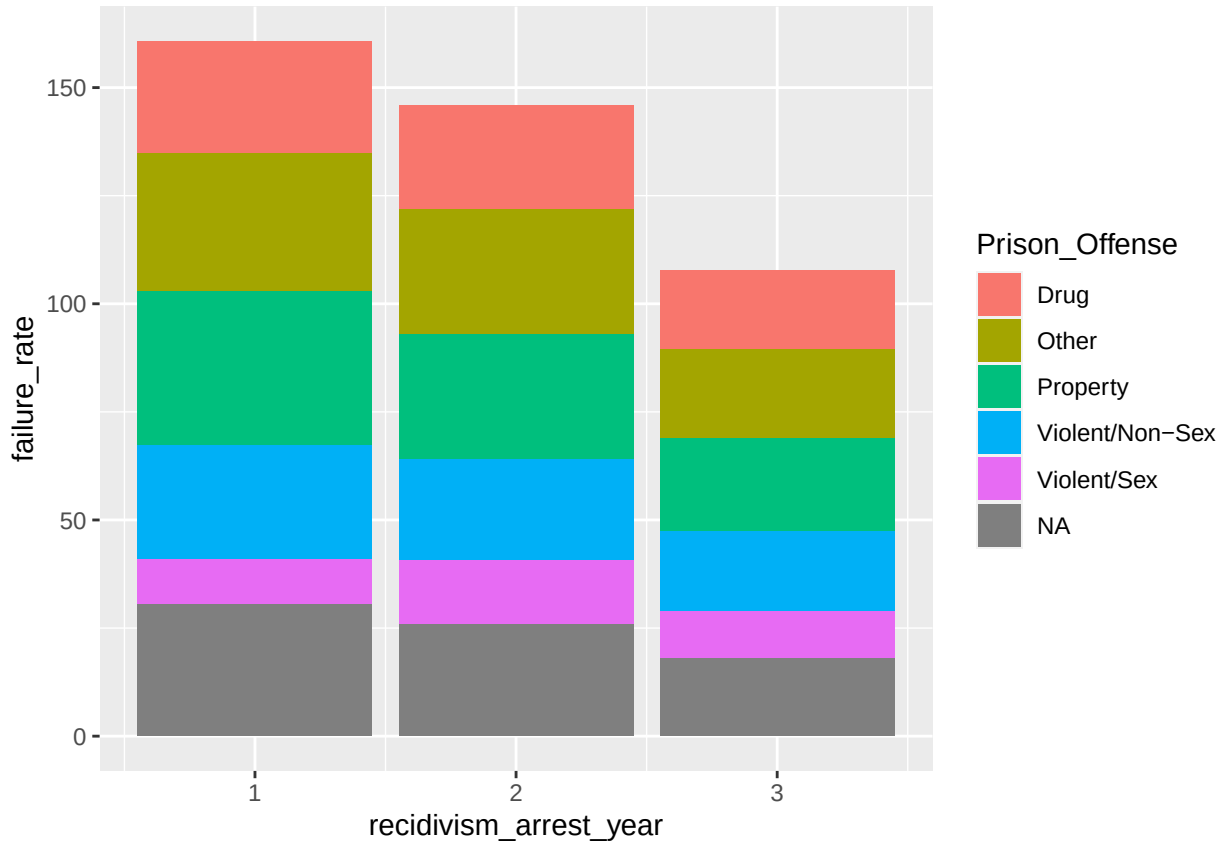
```
nij_fr_cr %>%
  filter(category == "Prison_Offense") %>%
  ggplot(aes(recidivism_arrest_year, failure_rate, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Prison_Offense")+
  geom_point()
```



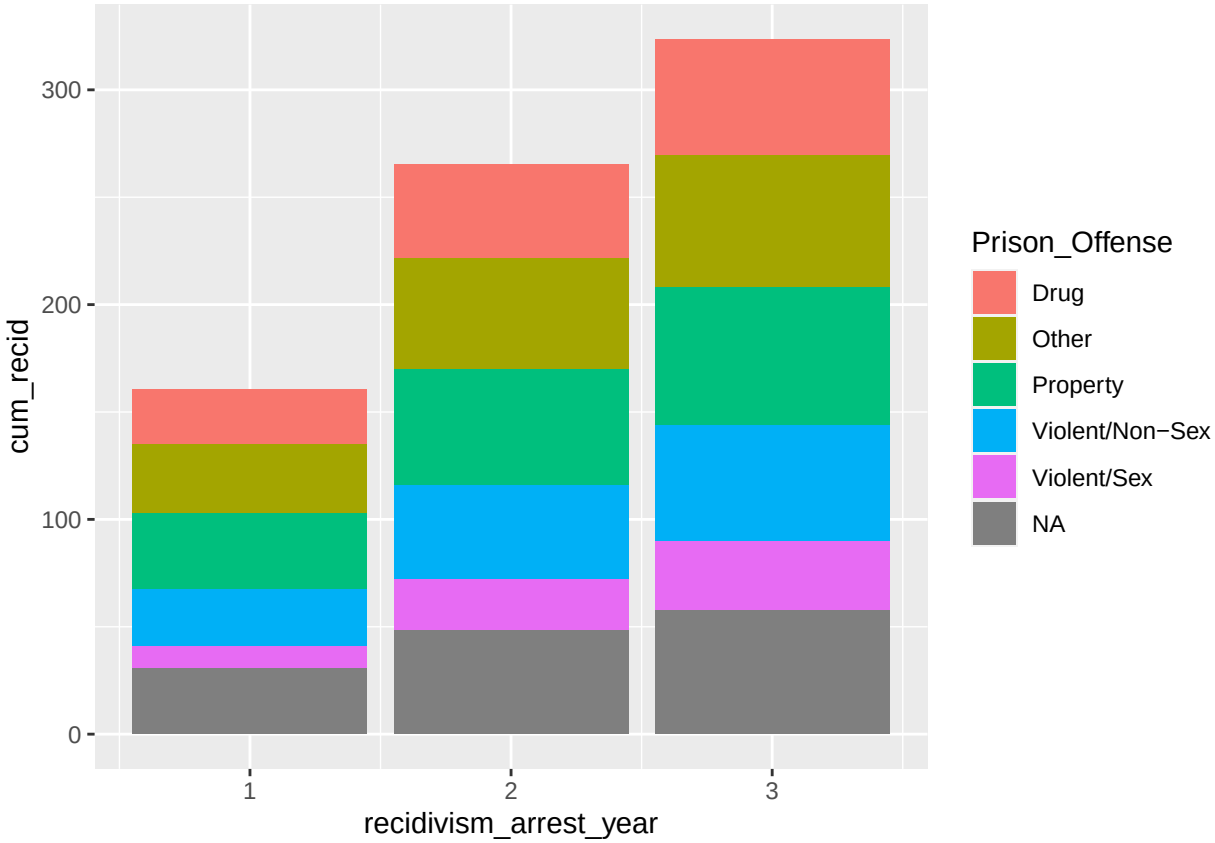

```
nij_fr_cr %>%
  filter(category == "Prison_Offense") %>%
  ggplot(aes(recidivism_arrest_year, cum_recid, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Prison_Offense")+
  geom_point()
```




```
#Stacked bar graph
nij_fr_cr %>%
  filter(category == "Prison_Offense") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = failure_rate), position="stack", stat=
  scale_fill_discrete(name="Prison_Offense")
```

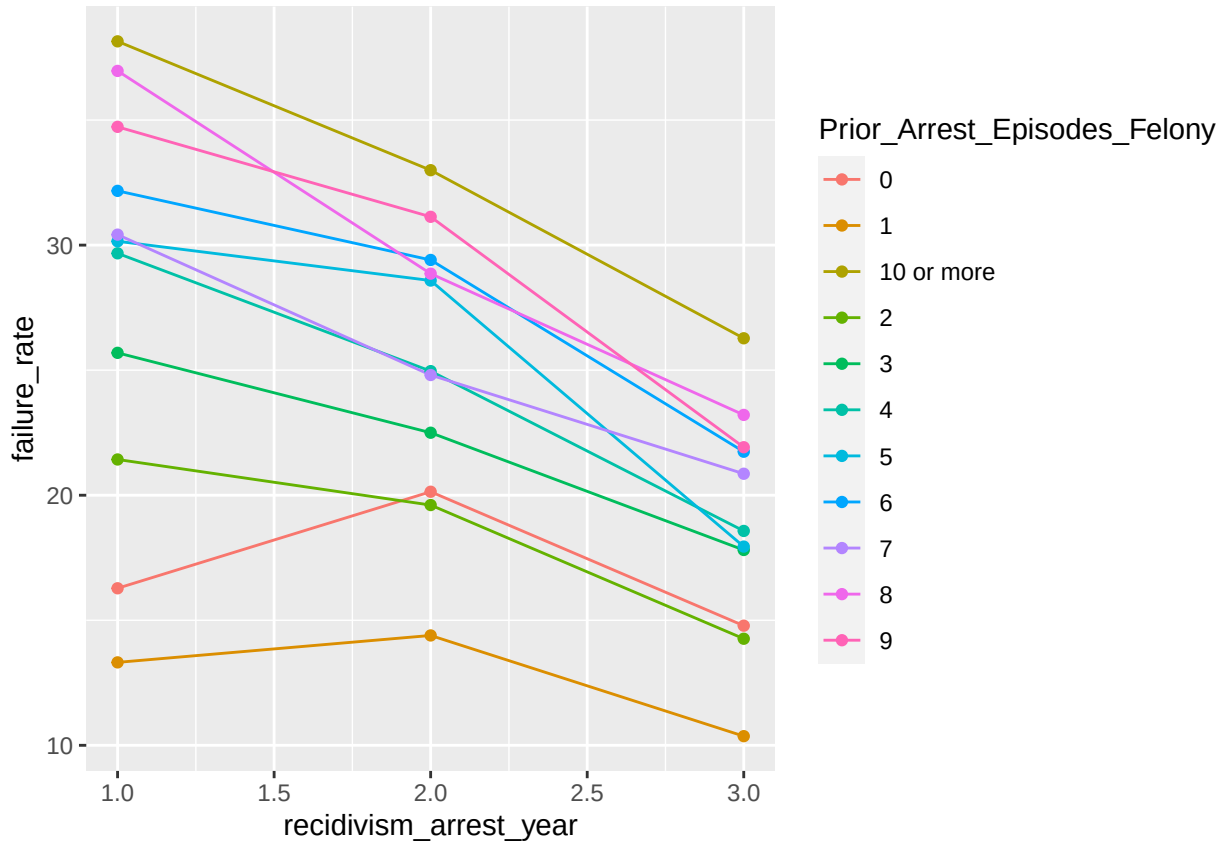



```
nij_fr_cr %>%
  filter(category == "Prison_Offense") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = cum_recid), position="stack", stat="id
  scale_fill_discrete(name="Prison_Offense")
```

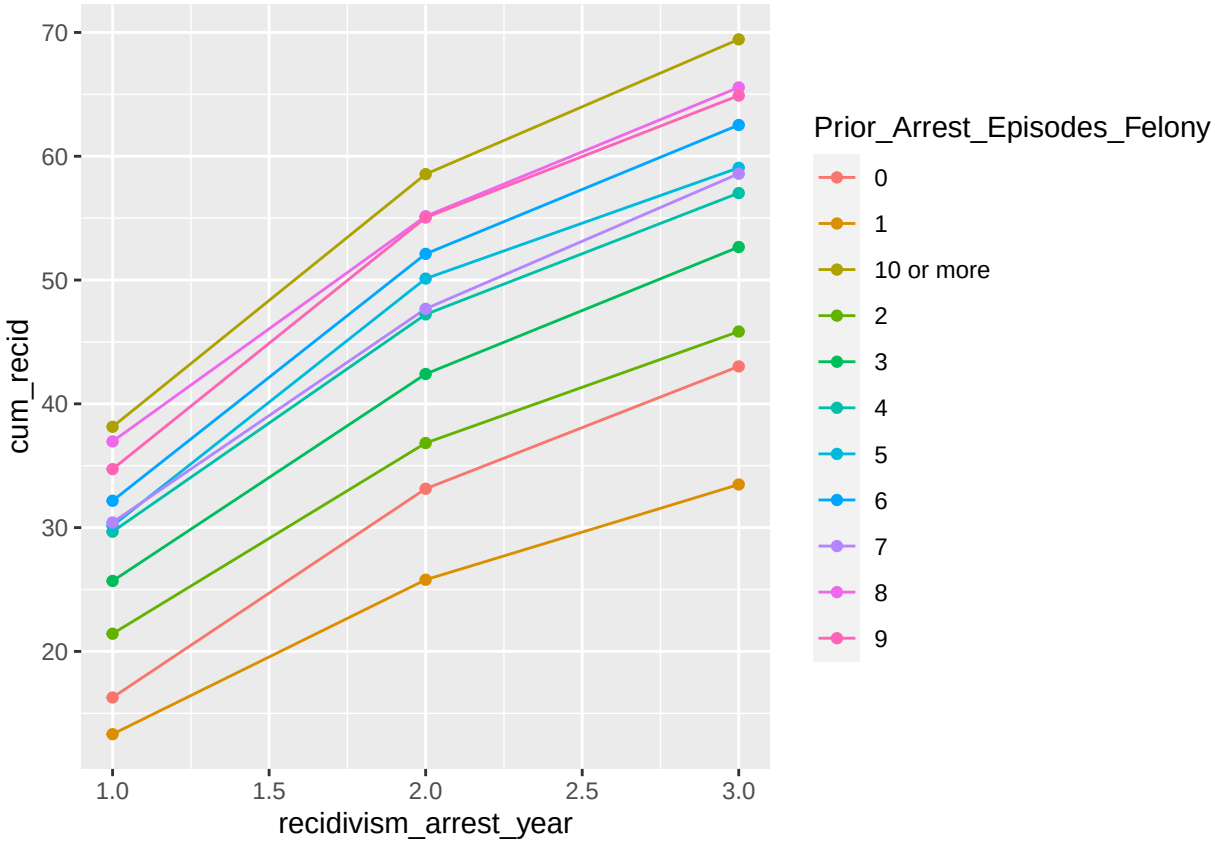


#Prior arrest episodes

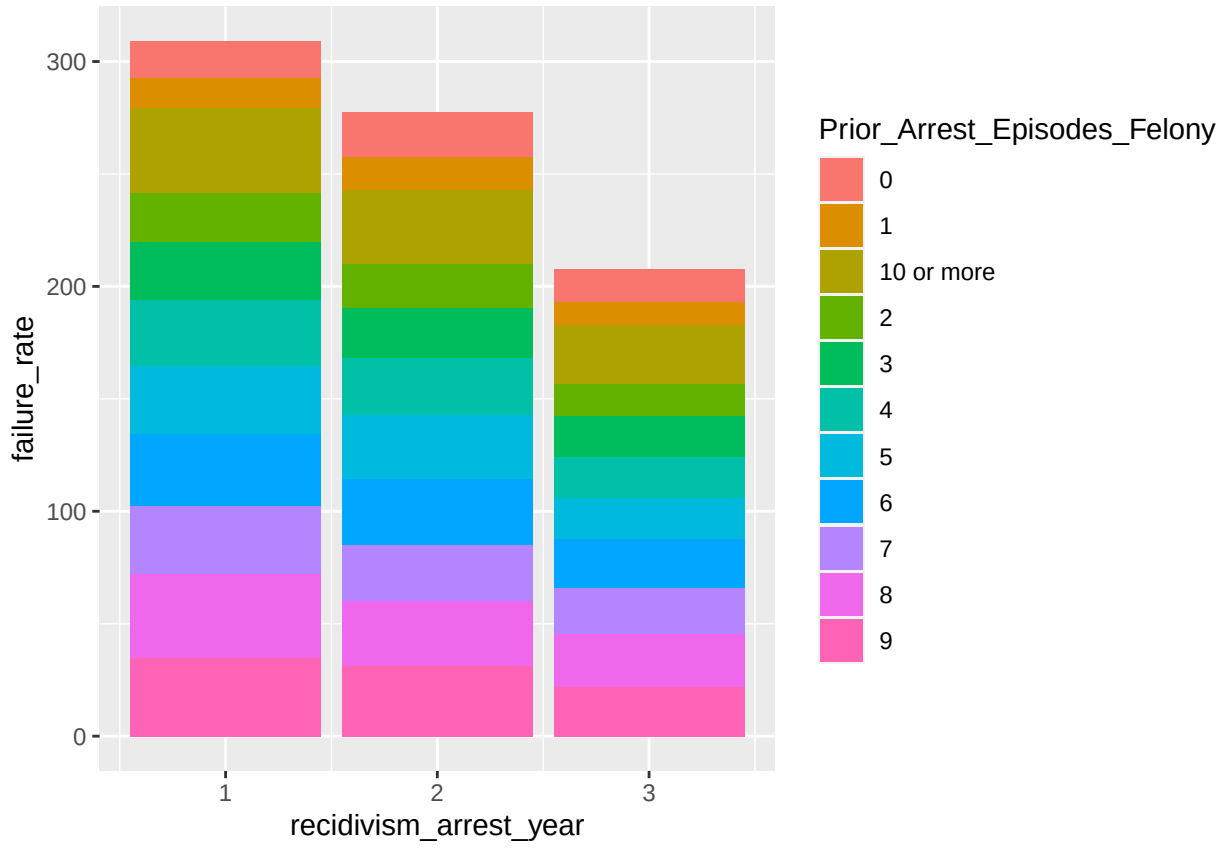
```
nij_fr_cr %>%
  filter(category == "Prior_Arrest_Episodes_Felony") %>%
  ggplot(aes(recidivism_arrest_year, failure_rate, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Prior_Arrest_Episodes_Felony")+
  geom_point()
```



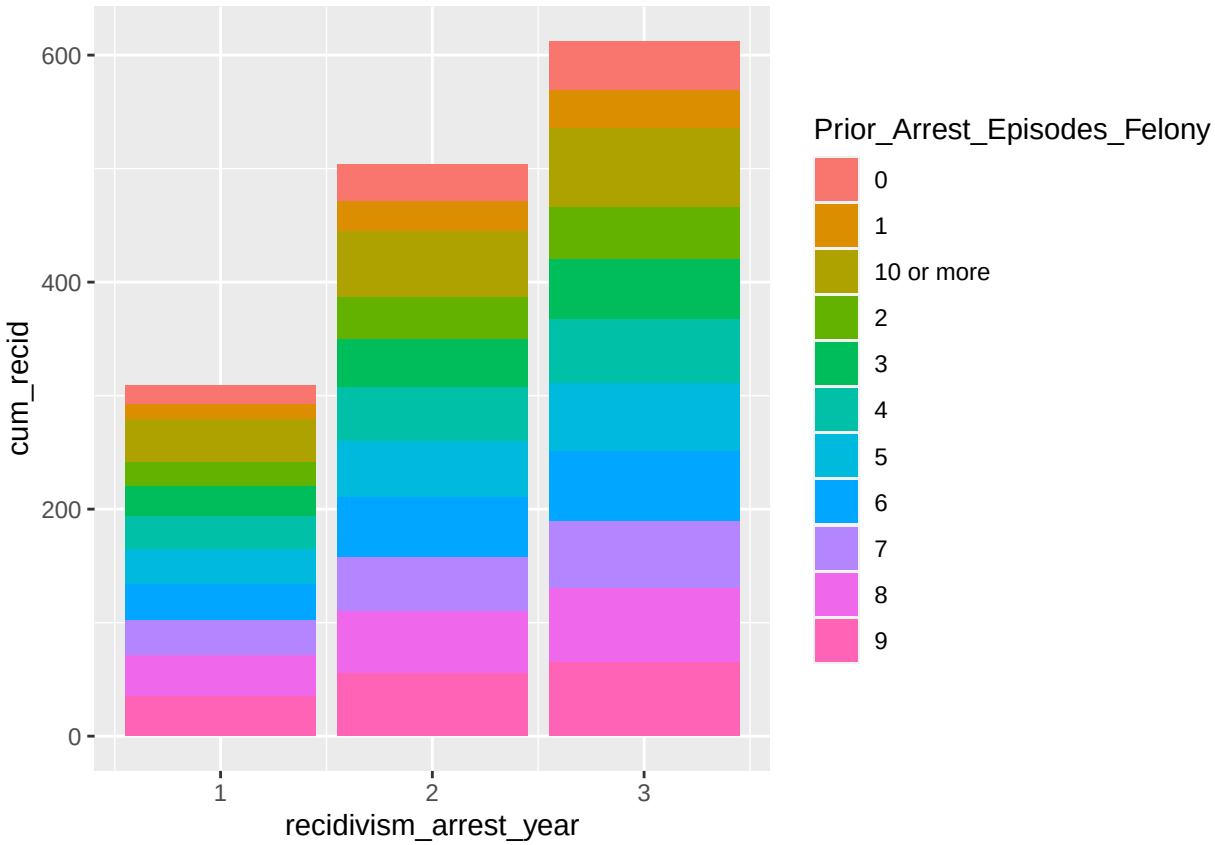

```
nij_fr_cr %>%
  filter(category == "Prior_Arrest_Episodes_Felony") %>%
  ggplot(aes(recidivism_arrest_year, cum_recid, color = attribute)) +
  geom_line() +
  scale_colour_discrete(name="Prior_Arrest_Episodes_Felony")+
  geom_point()
```




```
#Stacked bar graph
nij_fr_cr %>%
  filter(category == "Prior_Arrest_Episodes_Felony") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = failure_rate), position="stack", stat=
  scale_fill_discrete(name="Prior_Arrest_Episodes_Felony")
```




```
nij_fr_cr %>%
  filter(category == "Prior_Arrest_Episodes_Felony") %>%
  ggplot() +
  geom_bar(aes(fill = attribute, x = recidivism_arrest_year, y = cum_recid), position="stack", stat="id
  scale_fill_discrete(name="Prior_Arrest_Episodes_Felony")
```


600
400
200
0
cum_recid
1
2
3
recidivism_arrest_year
Prior_Arrest_Episodes_Felony
0
1
10 or more
2
3
4
5
6
7
8
9