

In []:

```
library(lubridate)
library(zoo)
library(foreign)
library(MonetDB.R)
library(DBI)
library(reshape2)
library(AICcmodavg)
library(bbmle)
library(stats4)
library(ggplot2)
library(stringr)
library(minpack.lm)
library(dplyr)
library(plyr)
library(tidyr)
library(jmv)
library(car)
library(ggplot2)
library(randomForest)
library(factoextra)
library(reshape2)
library("stringr")
library("ggbiplot")
library(effsize)
library(ggsci)
library(tidyverse)
```

In []:

```
IBB <- c(2,1,0,20) # 申告敬遠の数
PA <- c(367, 425, 175, 639) # 打席数
GP <- c(104, 106, 44, 155) # 試合数
# options(repr.plot.width=4, repr.plot.height=3)
```

In []:

```
batting_result<-read.delim("data_resh/batting_ohtani_resultOnly.tsv", sep='\t', header=T, stringsAs
Factor=FALSE)
pitching_result<-read.delim("data_resh/pitching_ohtani_resultOnly.tsv", sep='\t', header=T, strings
AsFactor=FALSE)
head(batting_result)
head(pitching_result)

pitching_split<-read.delim("data_resh/pitching_split.tsv", sep='\t', header=T, stringsAsFactor=FAL
SE)
head(pitching_split)
```

投手

月ごとの防御率

In [280]:

```
pitching_split %>%
  ggplot(aes(x=factor(month), y=ERA, color=factor(year))) + geom_point(size=4) + geom_line(aes(g
roup=factor(year)), size=2)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitl
e("era by month")

pitching_split %>%
  ggplot(aes(x=factor(month), y=IP / G, color=factor(year))) + geom_point(size=4) + geom_line(aes(
group=factor(year)), size=2)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitl
e("ip/g by month")
```

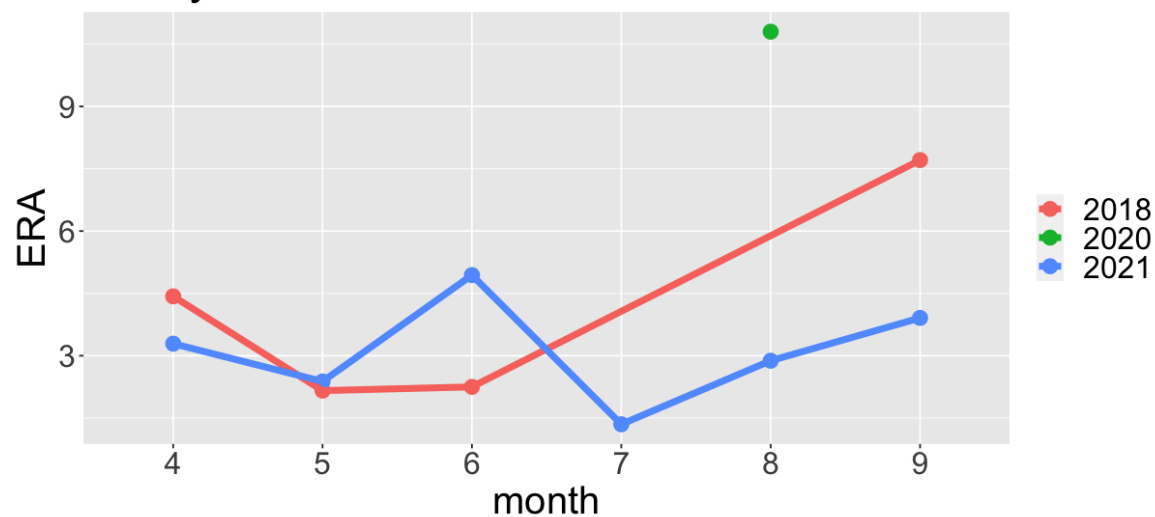
Warning message:

“Removed 1 rows containing missing values (geom_point).”

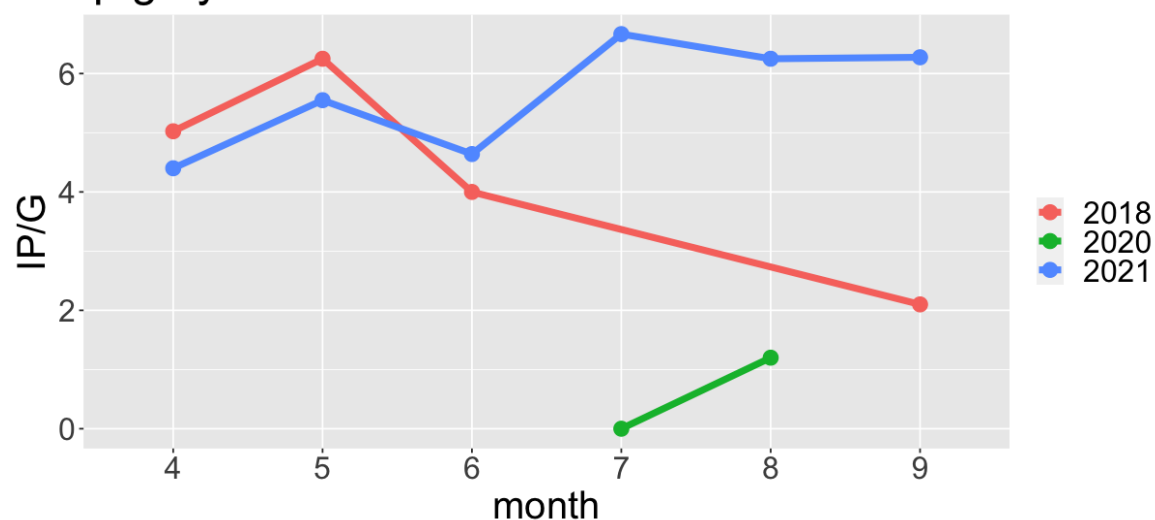
Warning message:

“Removed 1 row(s) containing missing values (geom_path).”

era by month



ip/g by month



In []:

```
all_pitches <- read.delim("data_res/pitching_ohtani.tsv", sep="\t", header=T, stringsAsFactor=FALSE)
```

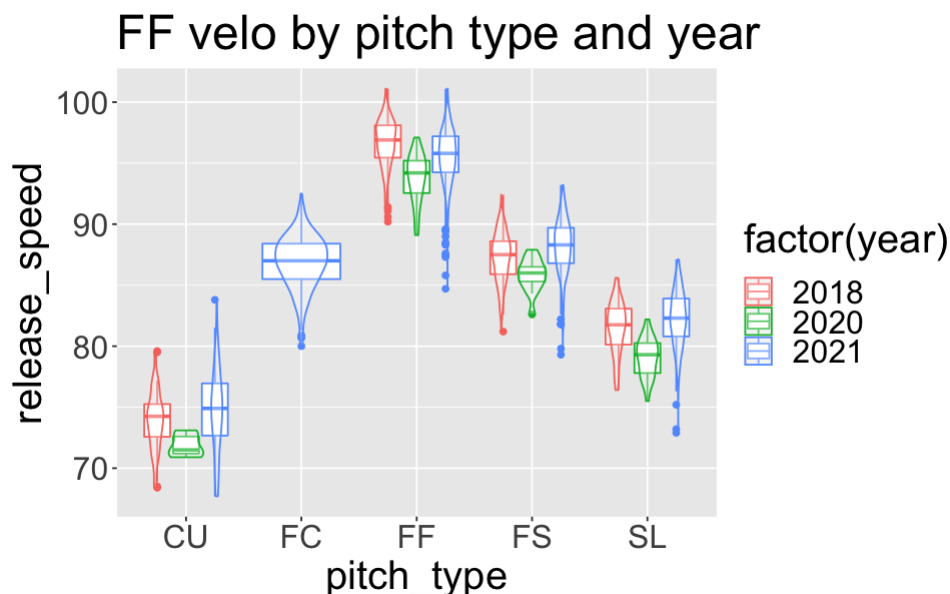
In [179]:

```
all_pitches$SL <- as.numeric(all_pitches$pitch_type == "SL")
all_pitches$FF <- as.numeric(all_pitches$pitch_type == "FF")
all_pitches$FC <- as.numeric(all_pitches$pitch_type == "FC")
all_pitches$FS <- as.numeric(all_pitches$pitch_type == "FS")
all_pitches$CU <- as.numeric(all_pitches$pitch_type == "CU")
```

球速の分布

In [474]:

```
options(repr.plot.width=8, repr.plot.height=5)
all_pitches %>%
  filter(pitch_type != "")%>%
  ggplot(aes(x=pitch_type, y=release_speed, color=factor(year))) + geom_boxplot() + theme(text = element_text(size = 24)) +
  ggtitle("FF velo by pitch type and year")+geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```



In [28]:

```
all_pitches$datenum <- as.numeric(gsub("-", "", all_pitches$game_date))
all_pitches$month_date <- all_pitches$datenum - round(all_pitches$datenum / 10000)*10000
all_pitches$month <- round(all_pitches$month_date / 100)
```

In [94]:

```
mean_ <- function(x){
  mm <- mean(x, na.rm = TRUE)
  ss <- sd(x, na.rm = TRUE)
  return(c(mm, ss))
}
```

In [465]:

```

velo_by_type_month <- aggregate(list("velo"=all_pitches$release_speed), list("year"=all_pitches$year, "month"=all_pitches$month, "type"=all_pitches$pitch_type), mean_)
velo_by_type_month$year <- factor(velo_by_type_month$year)
velo_by_type_month <- subset(velo_by_type_month, velo_by_type_month$type!="")

```

In [475]:

```

options(repr.plot.width=10, repr.plot.height=5)
all_pitches %>%
  filter(pitch_type == "FF")%>%
  ggplot(aes(x=factor(month), y=release_speed, color=factor(year))) + geom_boxplot() + ylim(90, 102) +
  geom_point(data = subset(velo_by_type_month, velo_by_type_month$type=="FF"),
             mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=4)+#, position=position_dodge(width = 0.7))
  geom_line(data = subset(velo_by_type_month, velo_by_type_month$type=="FF"),
            mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=2)+#, position=position_dodge(width = 0.7))
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("FF velo by year")+geom_violin(alpha=0.3, position=position_dodge(width=0.8))

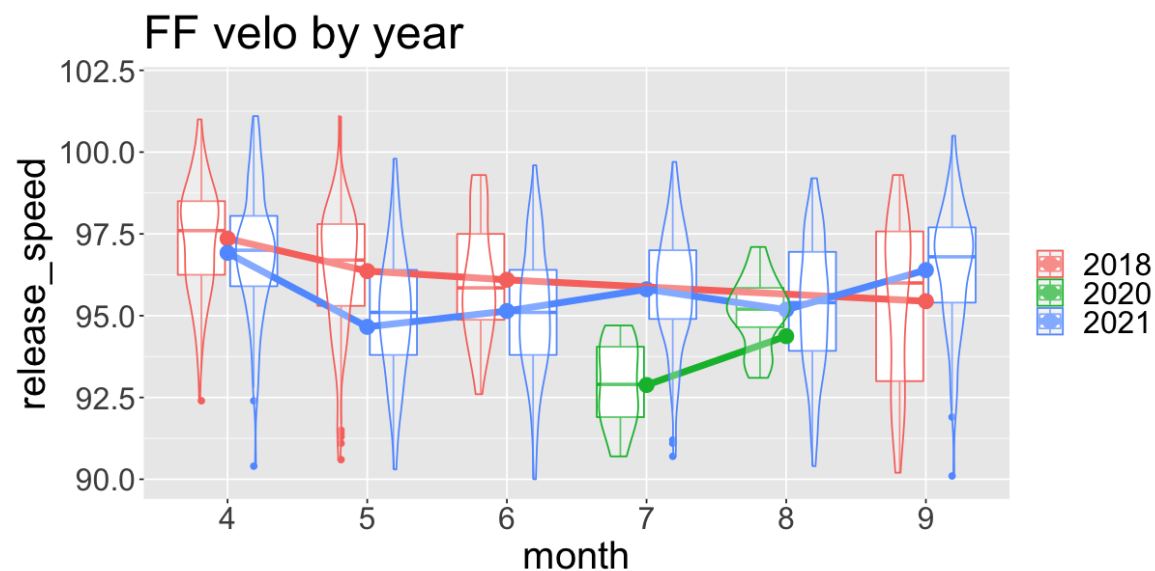
```

Warning message:

"Removed 17 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 17 rows containing non-finite values (stat_ydensity)."



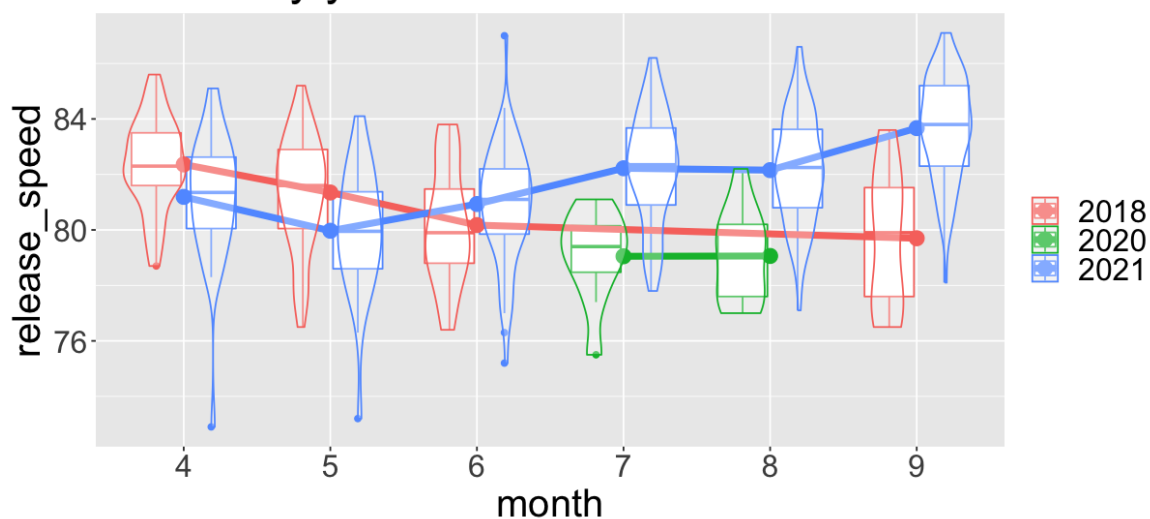
In [476]:

```

options(repr.plot.width=10, repr.plot.height=5)
all_pitches %>%
filter(pitch_type == "SL")%>%
ggplot(aes(x=factor(month), y=release_speed, color=factor(year))) + geom_boxplot()+ #ylim(90, 1
02) +
geom_point(data = subset(velo_by_type_month, velo_by_type_month$type=="SL"),
mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), s
ize=4)+#, position=position_dodge(width = 0.7))
geom_line(data = subset(velo_by_type_month, velo_by_type_month$type=="SL"),
mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), s
ize=2)+#, position=position_dodge(width = 0.7))
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitl
e("SL velo by year")+geom_violin(alpha=0.3, position=position_dodge(width=0.8))

```

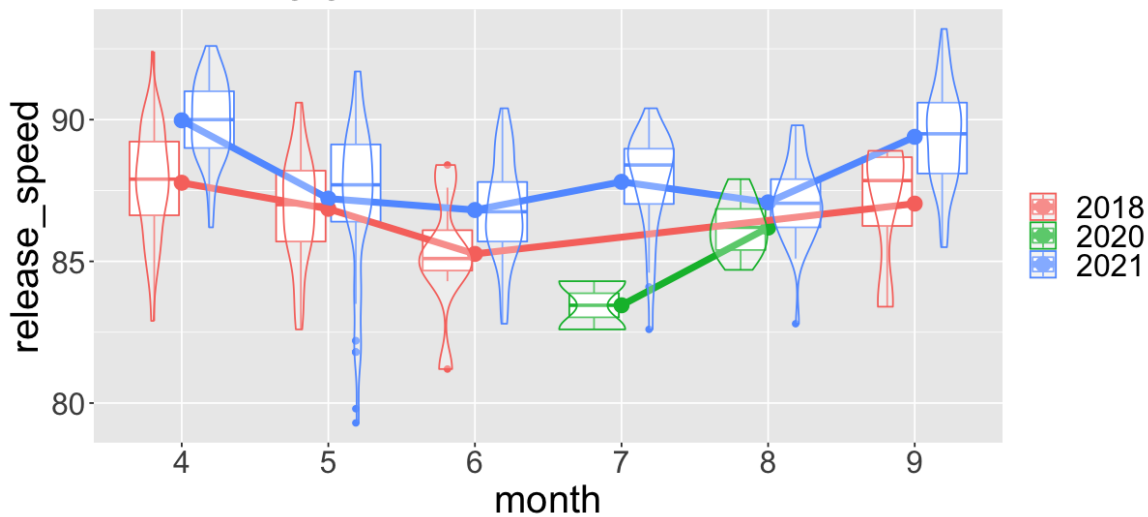
SL velo by year



In [477]:

```
options(repr.plot.width=10, repr.plot.height=5)
all_pitches %>%
  filter(pitch_type == "FS")%>%
  ggplot(aes(x=factor(month), y=release_speed, color=factor(year))) + geom_boxplot()+ #ylim(90, 102) +
  geom_point(data = subset(velo_by_type_month, velo_by_type_month$type=="FS"),
             mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=4)+#, position=position_dodge(width = 0.7))
  geom_line(data = subset(velo_by_type_month, velo_by_type_month$type=="FS"),
            mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=2)+#, position=position_dodge(width = 0.7))
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("FS velo by year")+geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

FS velo by year



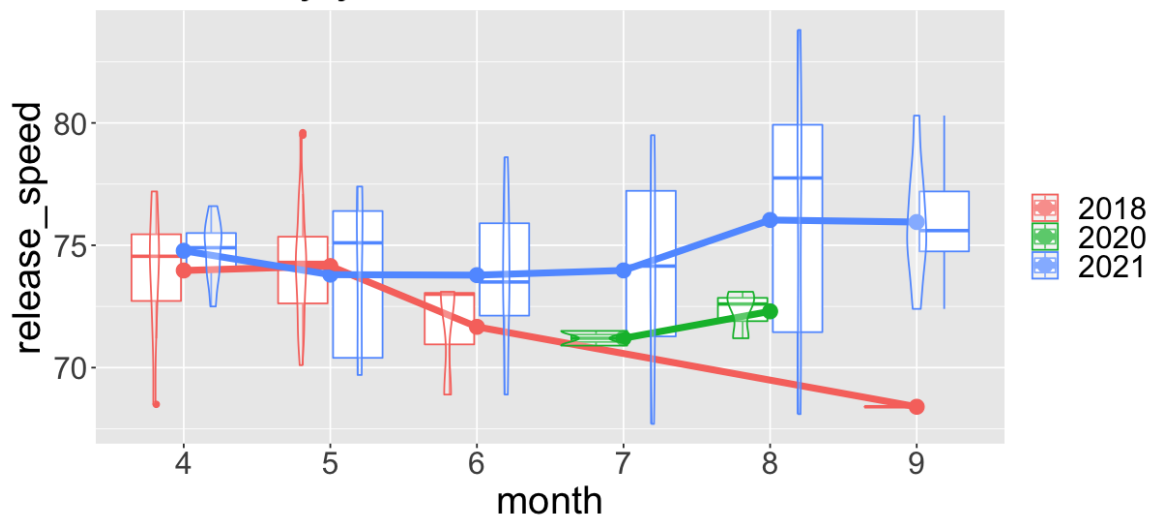
In [478]:

```
options(repr.plot.width=10, repr.plot.height=5)
all_pitches %>%
  filter(pitch_type == "CU") %>%
  ggplot(aes(x=factor(month), y=release_speed, color=factor(year))) + geom_boxplot() + #ylim(90, 102) +
  geom_point(data = subset(velo_by_type_month, velo_by_type_month$type=="CU"),
             mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=4) + #, position=position_dodge(width = 0.7))
  geom_line(data = subset(velo_by_type_month, velo_by_type_month$type=="CU"),
            mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=2) + #, position=position_dodge(width = 0.7))
  theme(text = element_text(size = 24)) + xlab("month") + theme(legend.title=element_blank()) + ggtitle("CU velo by year") + geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

Warning message:

"Groups with fewer than two data points have been dropped."

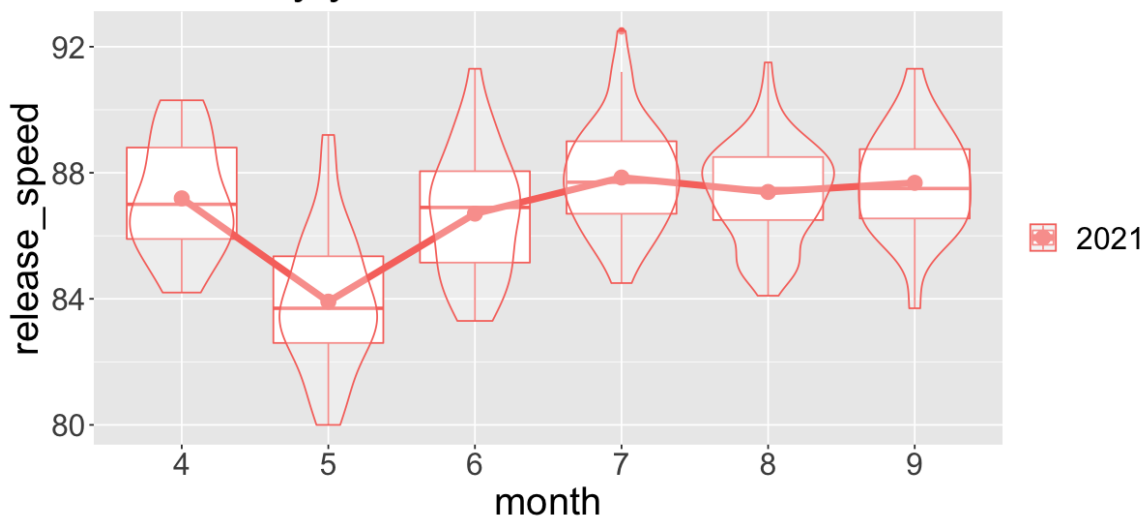
CU velo by year



In [479]:

```
options(repr.plot.width=10, repr.plot.height=5)
all_pitches %>%
  filter(pitch_type == "FC")%>%
  ggplot(aes(x=factor(month), y=release_speed, color=factor(year))) + geom_boxplot() + #ylim(90, 102) +
  geom_point(data = subset(velo_by_type_month, velo_by_type_month$type=="FC"),
             mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=4)+#, position=position_dodge(width = 0.7))
  geom_line(data = subset(velo_by_type_month, velo_by_type_month$type=="FC"),
            mapping = aes(x = factor(month), y = velo[,1], color=factor(year), group = factor(year)), size=2)+#, position=position_dodge(width = 0.7))
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) +
  ggtitle("FC velo by year")+geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

FC velo by year



In [223]:

```
velo_by_type_month_wide <- spread(velo_by_type_month, type, velo)
```

配球割合

In [184]:

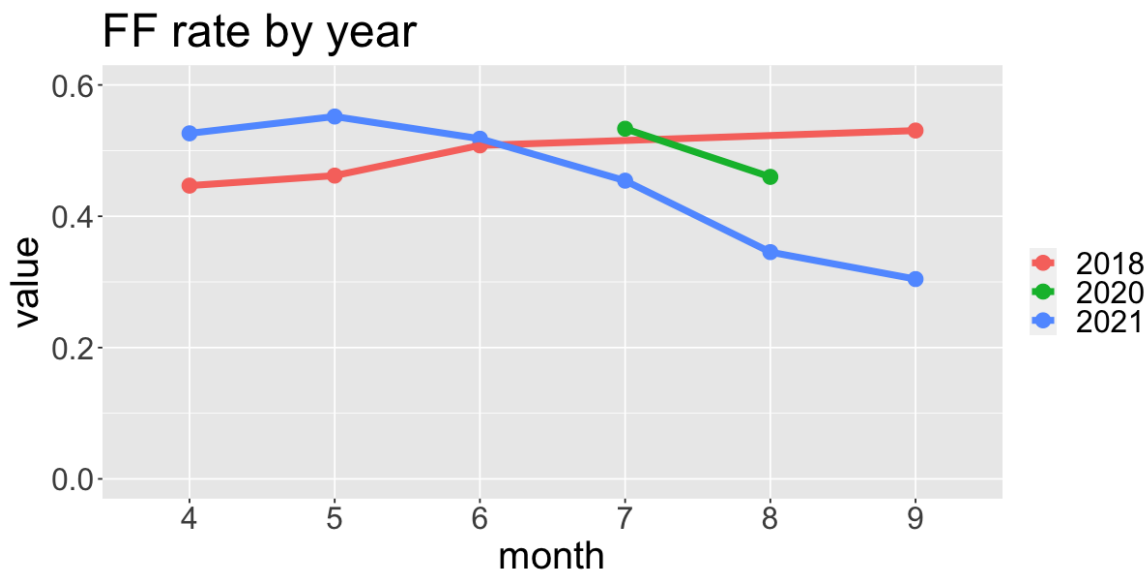
```
np_by_month <- aggregate(list("np"=all_pitches$isPitching, "nFF"=all_pitches$FF, "nFS"=all_pitches$FS, "nFC"=all_pitches$FC, "nSL"=all_pitches$SL, "nCU"=all_pitches$CU), list("year"=all_pitches$year, "month"=all_pitches$month), sum)
```


In [234]:

```
np_by_month$rFF <- np_by_month$nFF / np_by_month$np
np_by_month$rFS <- np_by_month$nFS / np_by_month$np
np_by_month$rFC <- np_by_month$nFC / np_by_month$np
np_by_month$rSL <- np_by_month$nSL / np_by_month$np
np_by_month$rCU <- np_by_month$nCU / np_by_month$np
np_by_month$rTS <- (np_by_month$nSL + np_by_month$nFC) / np_by_month$np
np_by_month_long <- gather(np_by_month, type, value, -year, -month)
np_by_month_long$type <- factor(np_by_month_long$type, levels=c("rFF", "rFS", "rTS", "rSL", "rFC", "rCU"))
```

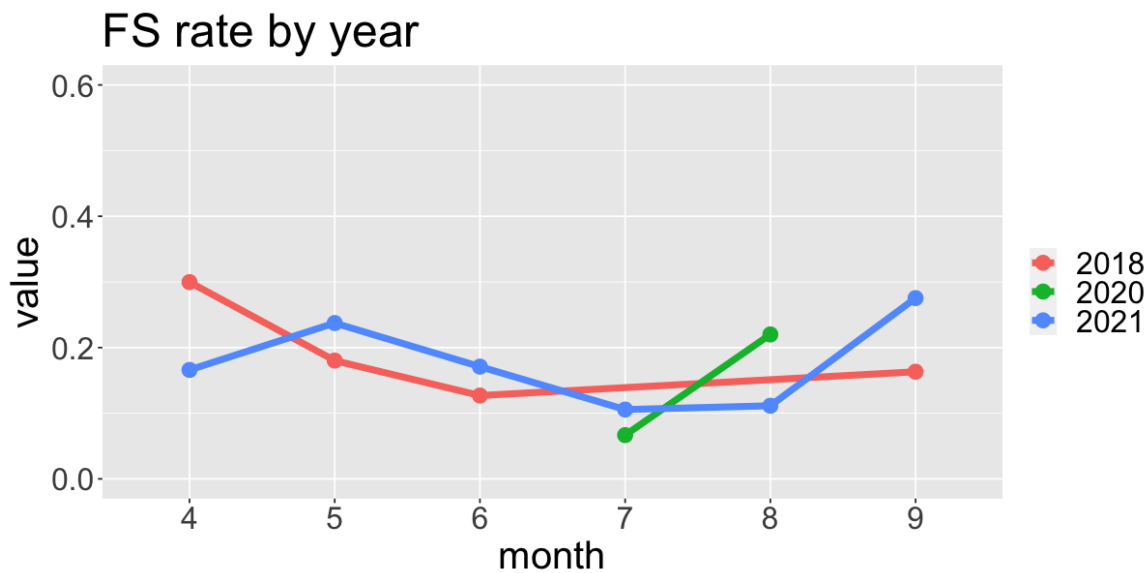
In [235]:

```
options(repr.plot.width=10, repr.plot.height=5)
#filter(type == "rFF" / type == "rFS" / type == "rSL" / type == "rFC" / type == "rCU")%>%
np_by_month_long %>%
filter(type == "rFF") %>%
ggplot(aes(x=factor(month), y=value, color=factor(year))) + geom_point(size=4) + geom_line(aes(
group=factor(year)), size=2)+
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle
("FF rate by year") + ylim(0,0.6)
```



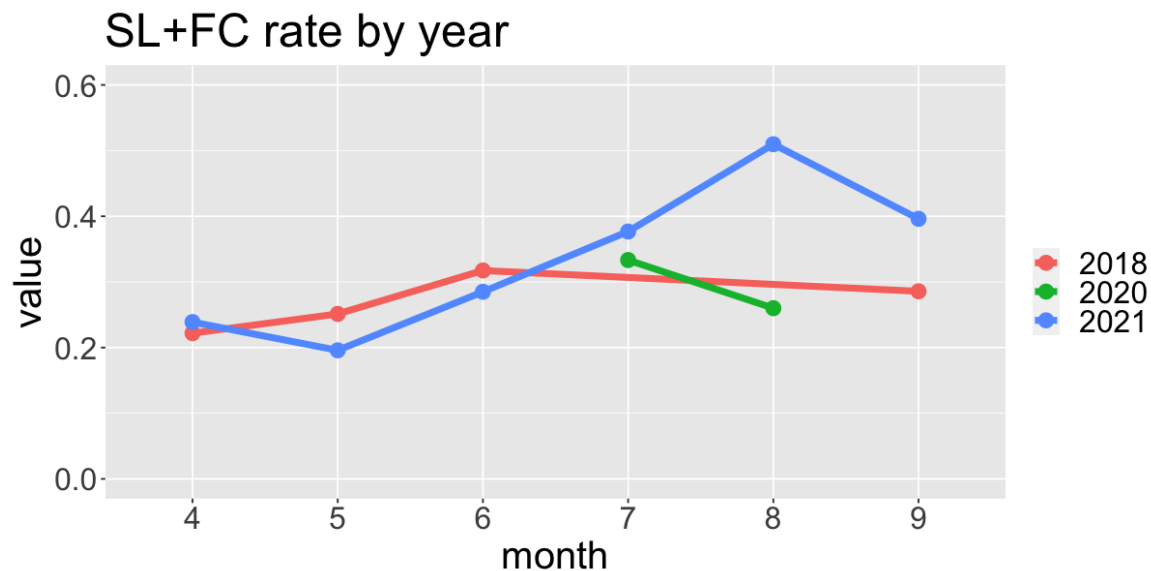
In [236]:

```
options(repr.plot.width=10, repr.plot.height=5)
#filter(type == "rFF" / type == "rFS" / type == "rSL" / type == "rFC" / type == "rCU")%>%
np_by_month_long %>%
filter(type == "rFS") %>%
ggplot(aes(x=factor(month), y=value, color=factor(year))) + geom_point(size=4) + geom_line(aes(
group=factor(year)), size=2)+
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle
("FS rate by year") + ylim(0,0.6)
```



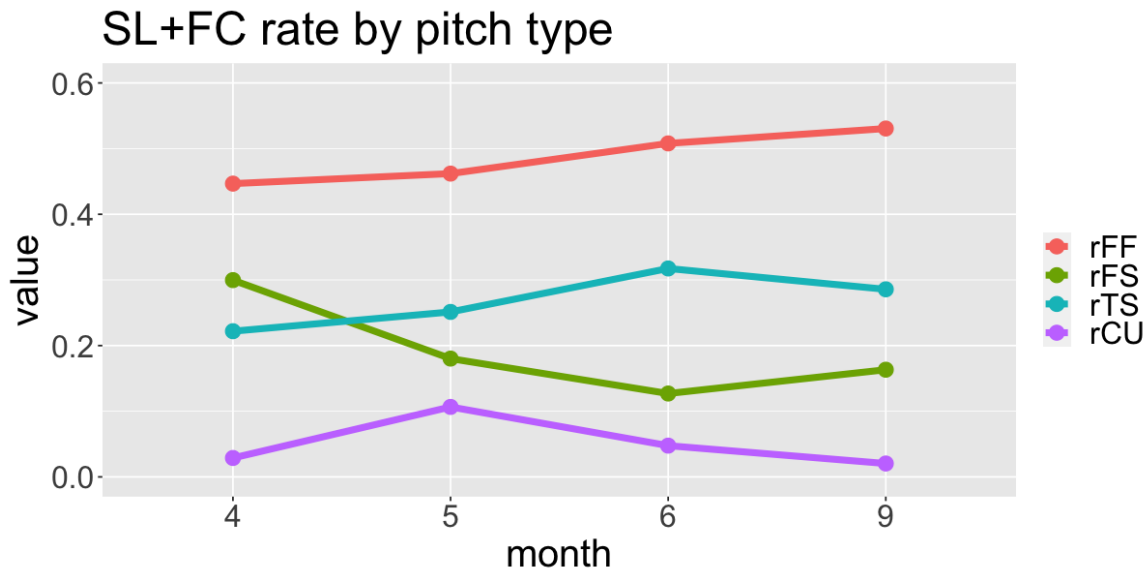
In [237]:

```
options(repr.plot.width=10, repr.plot.height=5)
np_by_month_long %>%
  filter(type == "rTS") %>%
  ggplot(aes(x=factor(month), y=value, color=factor(year))) + geom_point(size=4) + geom_line(aes(
    group=factor(year)), size=2)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle
e("SL+FC rate by year") + ylim(0,0.6)
```



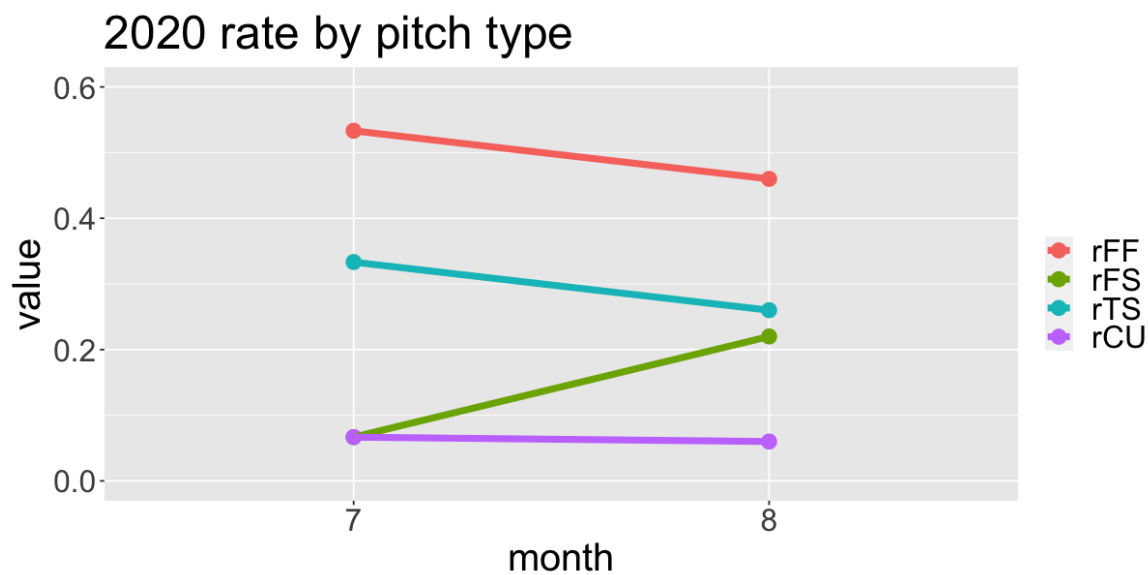
In [238]:

```
options(repr.plot.width=10, repr.plot.height=5)
np_by_month_long %>%
  filter(type == "rFF" | type == "rFS" | type == "rCU" | type == "rTS")%>%
  filter(year == 2018) %>%
  ggplot(aes(x=factor(month), y=value, color=type)) + geom_point(size=4) + geom_line(aes(group=t
ype), size=2)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle
("SL+FC rate by pitch type") + ylim(0,0.6)
```



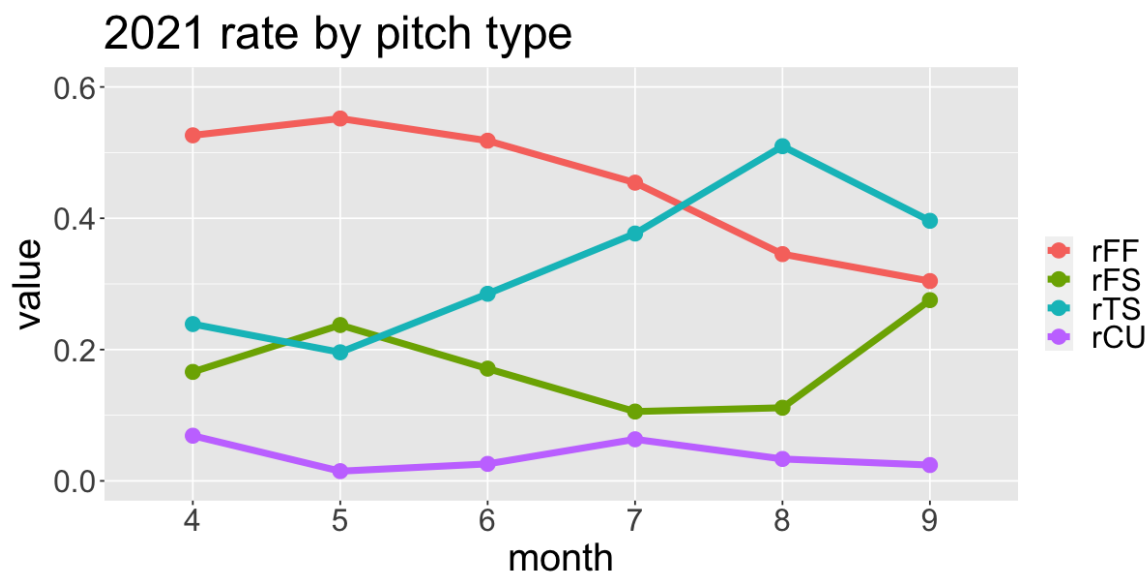
In [364]:

```
options(repr.plot.width=10, repr.plot.height=5)
np_by_month_long %>%
  filter(type == "rFF" | type == "rFS" | type == "rCU" | type == "rTS")%>%
  filter(year == 2020) %>%
  ggplot(aes(x=factor(month), y=value, color=type)) + geom_point(size=4) + geom_line(aes(group=type), size=2)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle(
    e("2020 rate by pitch type") + ylim(0,0.6)
```



In [365]:

```
options(repr.plot.width=10, repr.plot.height=5)
np_by_month_long %>%
  filter(type == "rFF" | type == "rFS" | type == "rCU" | type == "rTS")%>%
  filter(year == 2021) %>%
  ggplot(aes(x=factor(month), y=value, color=type)) + geom_point(size=4) + geom_line(aes(group=t
ype), size=2)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtil
e("2021 rate by pitch type") + ylim(0,0.6)
```



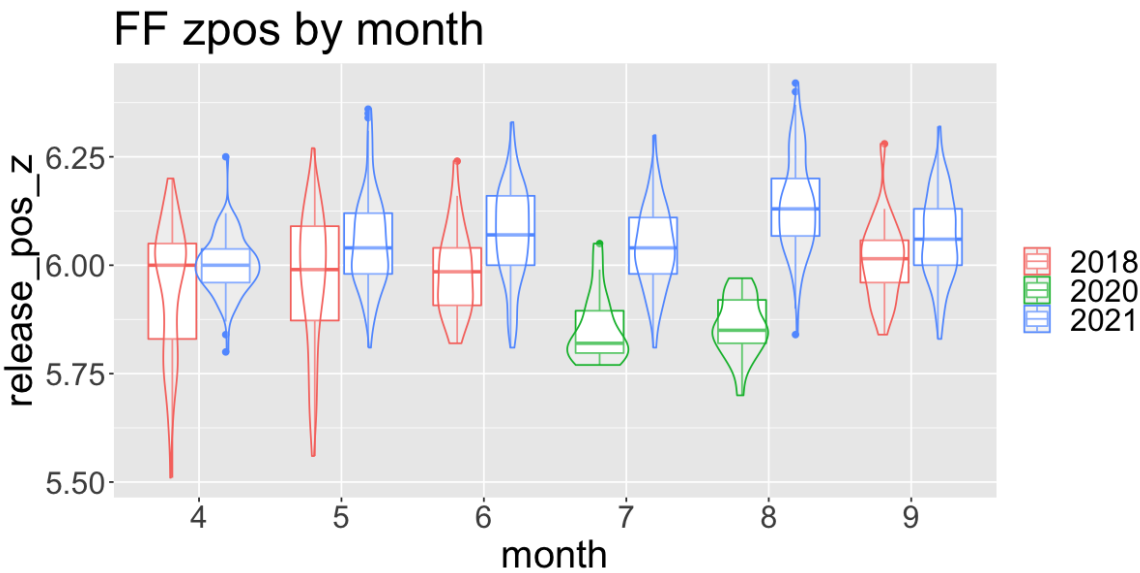
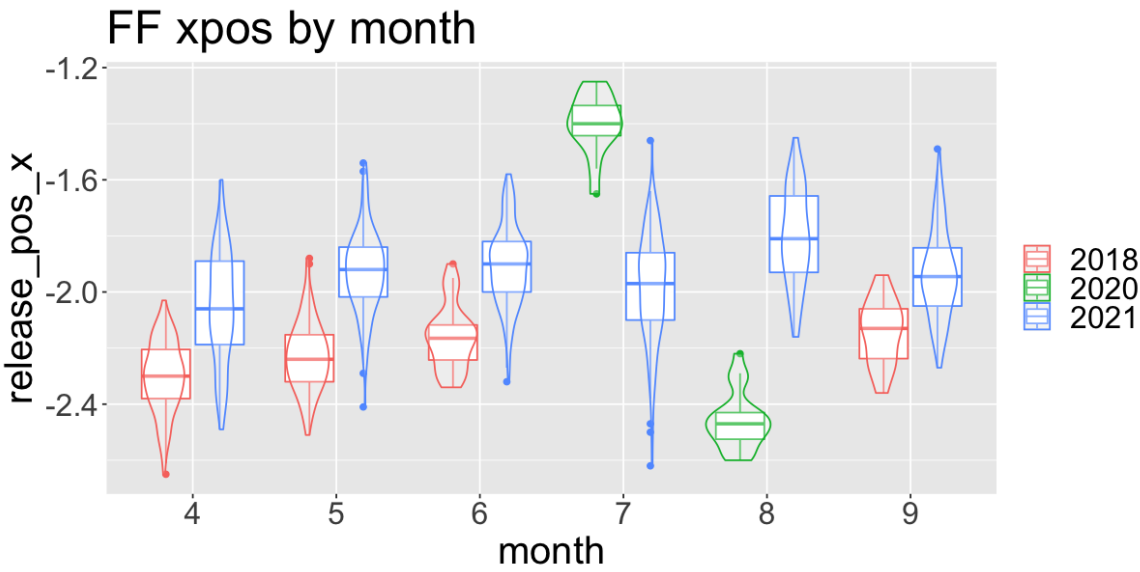
リリースポイント

In [481]:

```
options(repr.plot.width=10, repr.plot.height=5)
all_pitches %>%
  filter(pitch_type == "FF")%>%
  ggplot(aes(x=factor(month), y=release_pos_x, color=factor(year))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("FF xpos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))

all_pitches %>%
  filter(pitch_type == "FF")%>%
  ggplot(aes(x=factor(month), y=release_pos_y, color=factor(year))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("FF ypos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))

all_pitches %>%
  filter(pitch_type == "FF")%>%
  ggplot(aes(x=factor(month), y=release_pos_z, color=factor(year))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("FF zpos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```



In [482]:

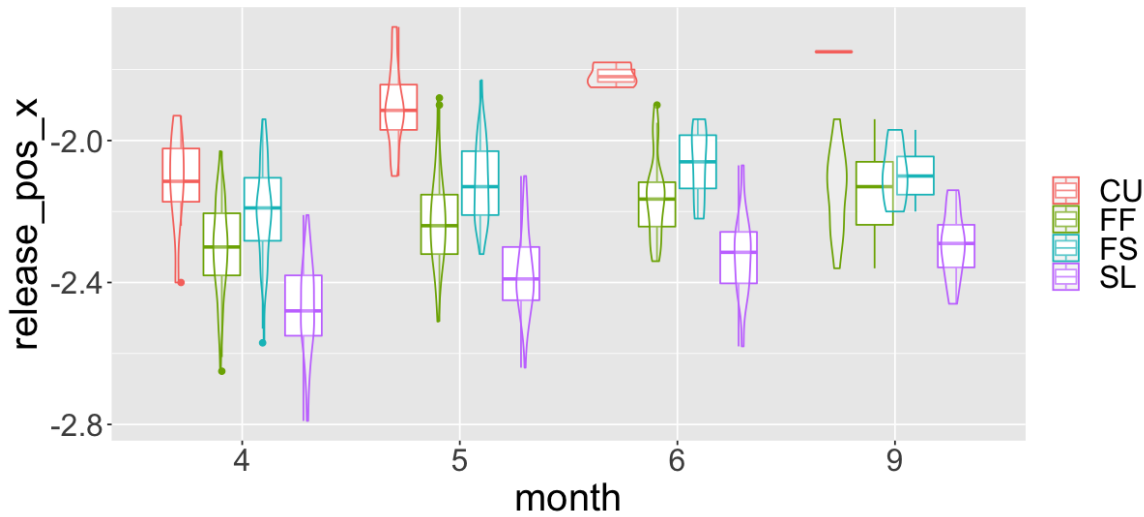
```
all_pitches %>%
  filter(year==2018)%>%
  filter(pitch_type != "")%>%
  ggplot(aes(x=factor(month), y=release_pos_x, color=factor(pitch_type))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("2018 xpos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))

all_pitches %>%
  filter(year==2018)%>%
  filter(pitch_type != "")%>%
  ggplot(aes(x=factor(month), y=release_pos_y, color=factor(pitch_type))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("2018 ypos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))

all_pitches %>%
  filter(year==2018)%>%
  filter(pitch_type != "")%>%
  ggplot(aes(x=factor(month), y=release_pos_z, color=factor(pitch_type))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("2018 zpos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

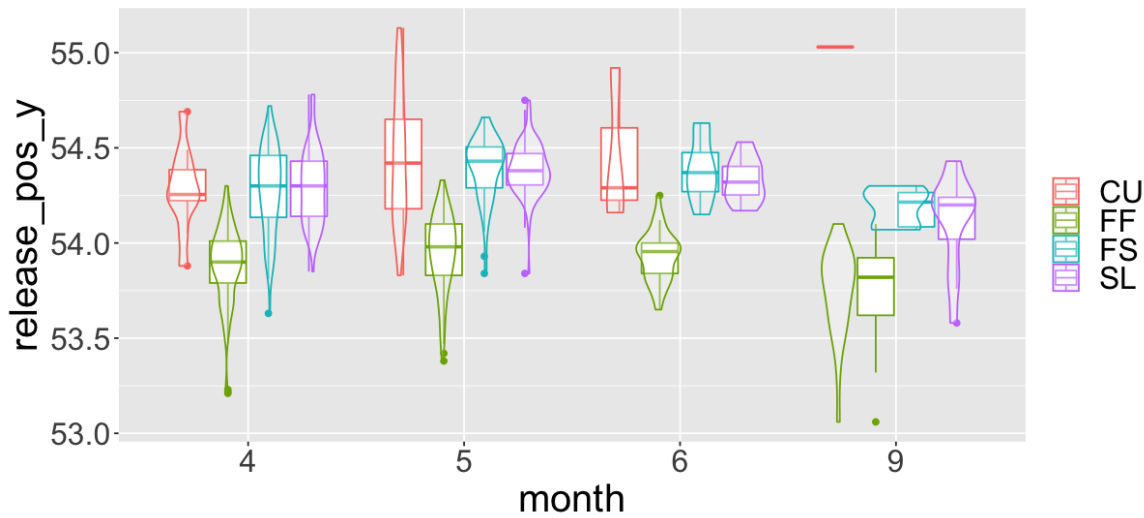
Warning message:
"Groups with fewer than two data points have been dropped."
Warning message:
"Groups with fewer than two data points have been dropped."

2018 xpos by month

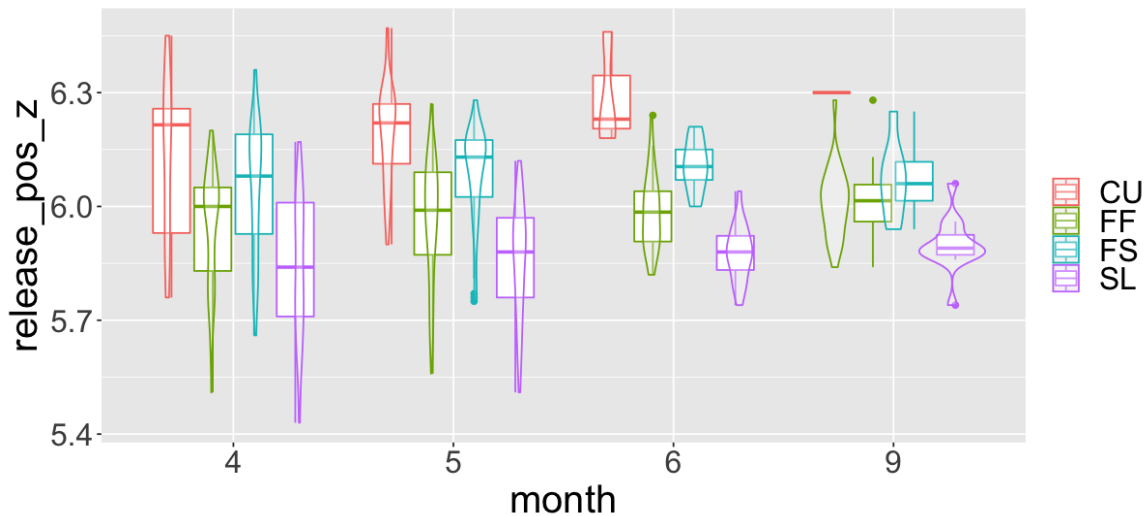


Warning message:
"Groups with fewer than two data points have been dropped."

2018 ypos by month



2018 zpos by month



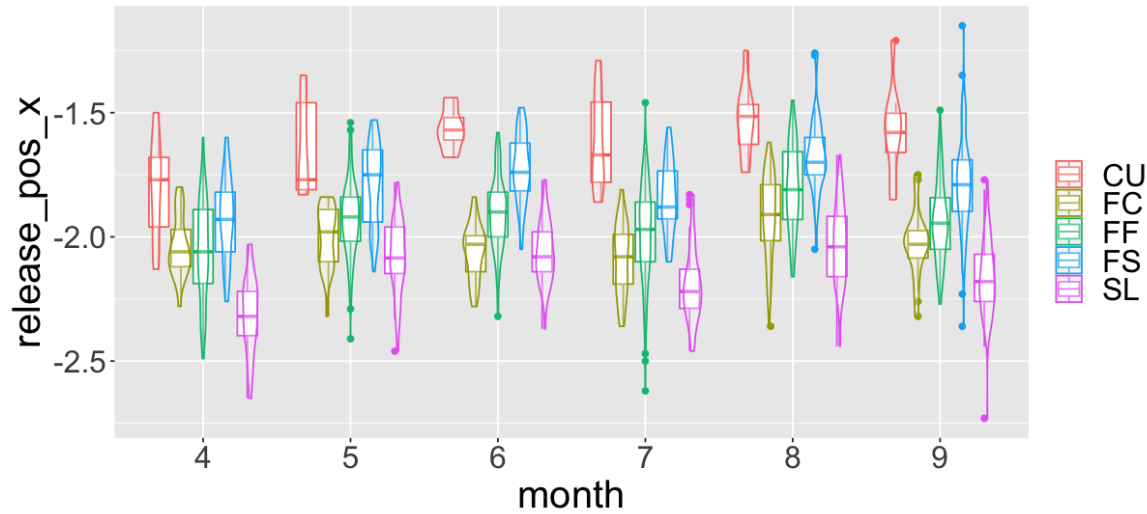
In [483]:

```
all_pitches %>%
  filter(year==2021)%>%
  ggplot(aes(x=factor(month), y=release_pos_x, color=factor(pitch_type))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("2021 xpos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))

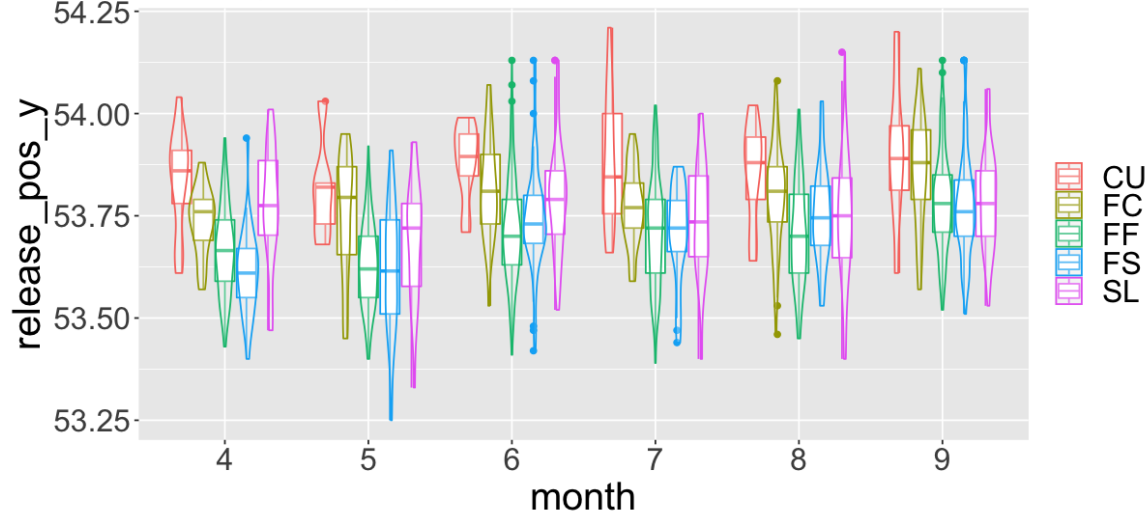
all_pitches %>%
  filter(year==2021)%>%
  ggplot(aes(x=factor(month), y=release_pos_y, color=factor(pitch_type))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("2021 ypos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))

all_pitches %>%
  filter(year==2021)%>%
  ggplot(aes(x=factor(month), y=release_pos_z, color=factor(pitch_type))) + geom_boxplot()+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("2021 zpos by month")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

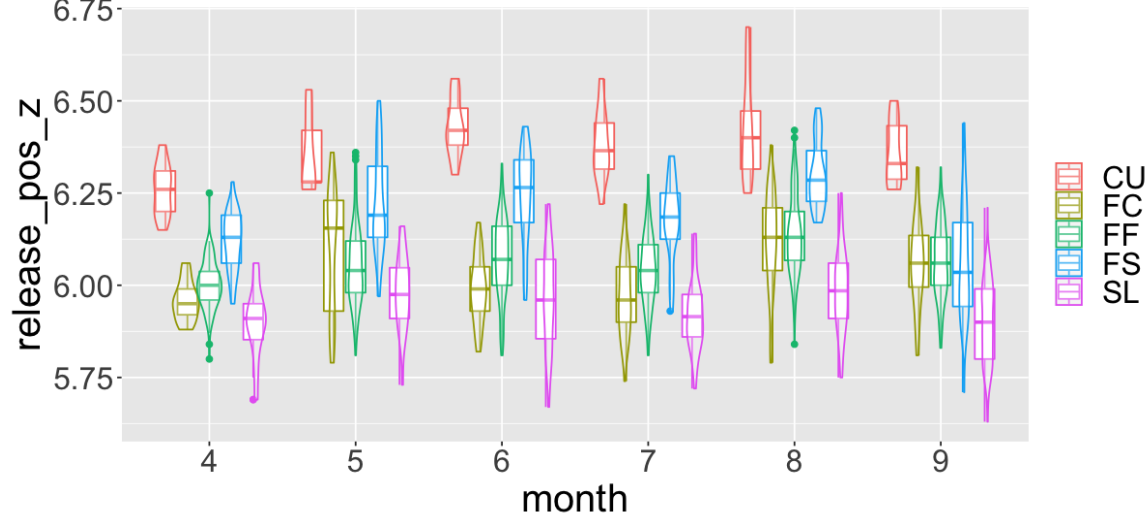
2021 xpos by month



2021 ypos by month



2021 zpos by month



リリースポイント分散

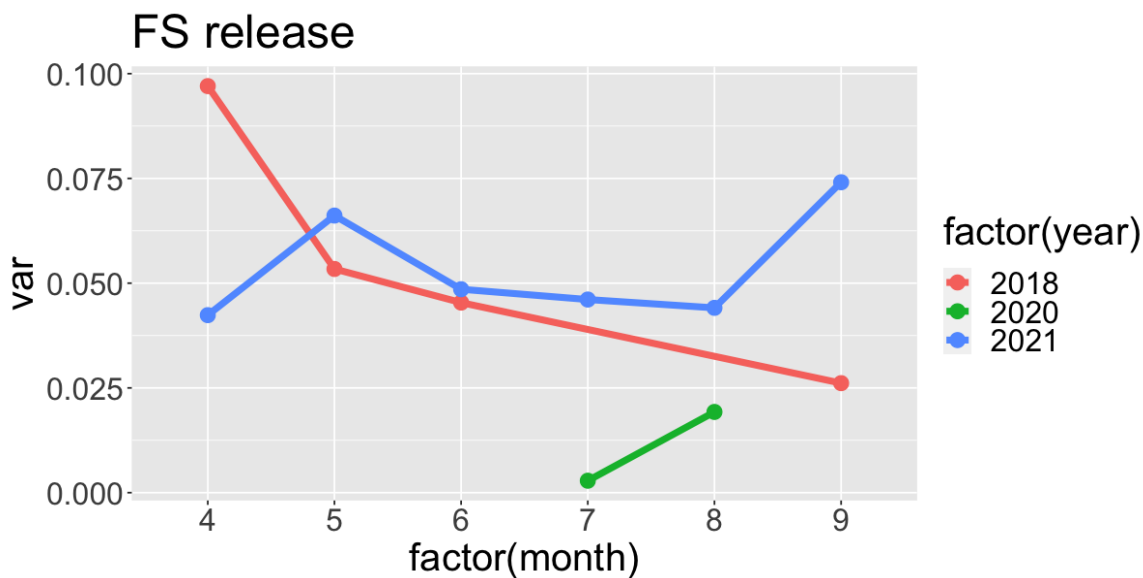
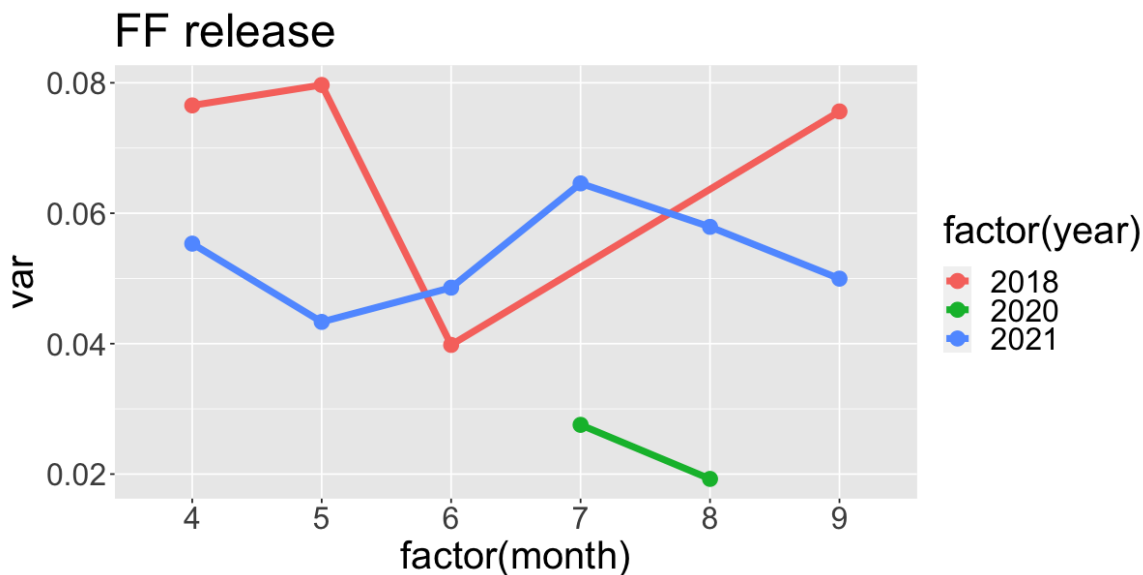
In [484]:

```
release_pos <- aggregate(list("pos_x"=all_pitches$release_pos_x, "pos_y"=all_pitches$release_pos_y, "pos_z"=all_pitches$release_pos_z), list("year"=all_pitches$year, "month"=all_pitches$month, "type"=all_pitches$pitch_type), mean_)
release_pos$var <- release_pos$pos_x[,2]*release_pos$pos_x[,2]+
release_pos$pos_y[,2]*release_pos$pos_y[,2]+
release_pos$pos_z[,2]*release_pos$pos_z[,2]
```

In [363]:

```
options(repr.plot.width=10, repr.plot.height=5)
release_pos %>%
filter(type == "FF")%>%
ggplot(aes(x=factor(month), y=var, color=factor(year))) + geom_point(size=4)+geom_line(size=2,
aes(group=factor(year)))+ theme(text = element_text(size = 24))+
ggtitle("FF release")

release_pos %>%
filter(type == "FS")%>%
ggplot(aes(x=factor(month), y=var, color=factor(year))) + geom_point(size=4)+geom_line(size=2,
aes(group=factor(year)))+ theme(text = element_text(size = 24))+
ggtitle("FS release")
```



In []:

野手

In [307]:

```
all_bats <- read.delim("data_res/batting_ohtani.tsv", sep="\t", header=T, stringsAsFactor=FALSE)
all_bats$datenum <- as.numeric(gsub("-", "", all_bats$game_date))
all_bats$month_date <- all_bats$datenum - round(all_bats$datenum / 10000)*10000
all_bats$month <- round(all_bats$month_date / 100)
```

In [488]:

```
all_bats$isBarrel <- as.numeric(all_bats$launch_speed_angle == 6)

exit_by_type_month <- aggregate(list("barrel_rate"=all_bats$isBarrel, "exit_velo"=all_bats$launch_s
peed, "launch_angle"=all_bats$launch_angle), list("year"=all_bats$year, "month"=all_bats$month),
mean_)
#head(exit_by_type_month)

pitch_count <- aggregate(list("np"=all_pitches$isPitching), list("year"=all_pitches$year, "month"=all
_pitches$month), sum)
pitch_count_b <- aggregate(list("npb"=all_bats$isPitching!=1), list("year"=all_bats$year, "month"=al
l_bats$month), sum)
#head(pitch_count)
#head(pitch_count_b)

#merge(pitch_count, pitch_count_b)
pitch_and_bat <- merge(exit_by_type_month, merge(pitch_count, pitch_count_b))

#head(pitch_and_bat)
```

打球速度&角度

In [489]:

```
options(repr.plot.width=10, repr.plot.height=8)
all_bats %>%
  ggplot(aes(x=factor(month), y=launch_speed, color=factor(year))) + geom_boxplot()+
  geom_point(data = exit_by_type_month,
             mapping = aes(x = factor(month), y = exit_velo[,1], color=factor(year), group = factor(year)),
             size=4)+#, position=position_dodge(width = 0.7))
  geom_line(data = exit_by_type_month,
            mapping = aes(x = factor(month), y = exit_velo[,1], color=factor(year), group = factor(year)),
            size=2)+#, position=position_dodge(width = 0.7))
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle
  e("exit velo by year")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

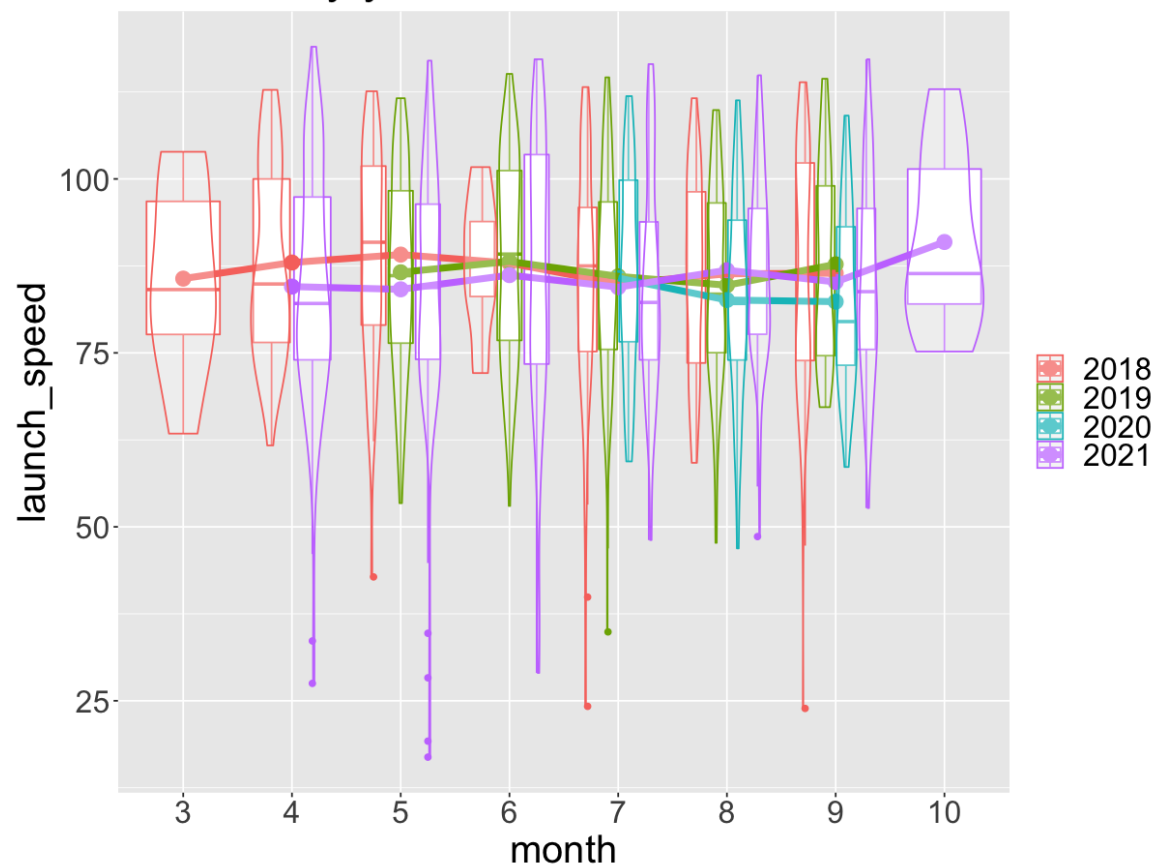
Warning message:

"Removed 4745 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 4745 rows containing non-finite values (stat_ydensity)."

exit velo by year



In [490]:

```
options(repr.plot.width=10, repr.plot.height=8)
all_bats %>%
  ggplot(aes(x=factor(month), y=launch_angle, color=factor(year))) + geom_boxplot()+
  geom_point(data = exit_by_type_month,
             mapping = aes(x = factor(month), y = launch_angle[,1], color=factor(year), group = factor
                           (year)), size=4)+#, position=position_dodge(width = 0.7))
  geom_line(data = exit_by_type_month,
            mapping = aes(x = factor(month), y = launch_angle[,1], color=factor(year), group = factor
                          (year)), size=2)+#, position=position_dodge(width = 0.7))
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle
  e("launch angle by year")+
  geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

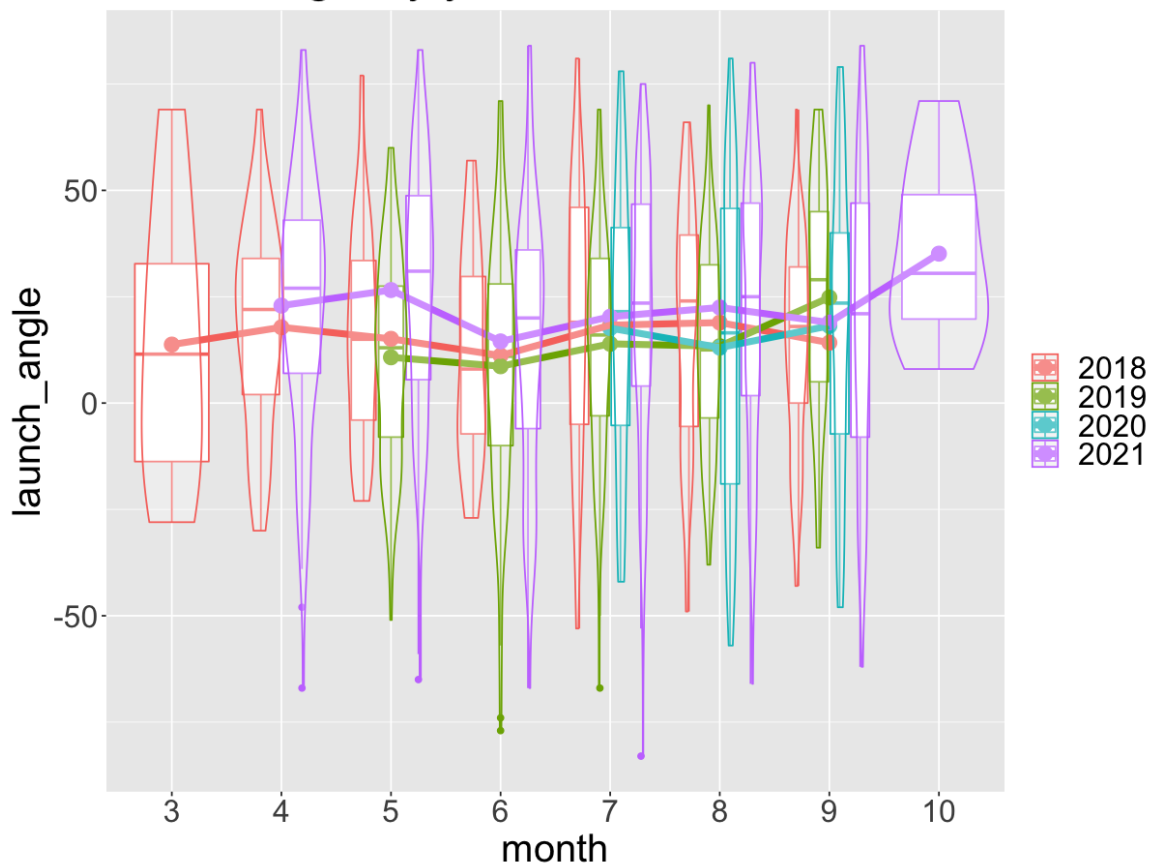
Warning message:

"Removed 4745 rows containing non-finite values (stat_boxplot)."

Warning message:

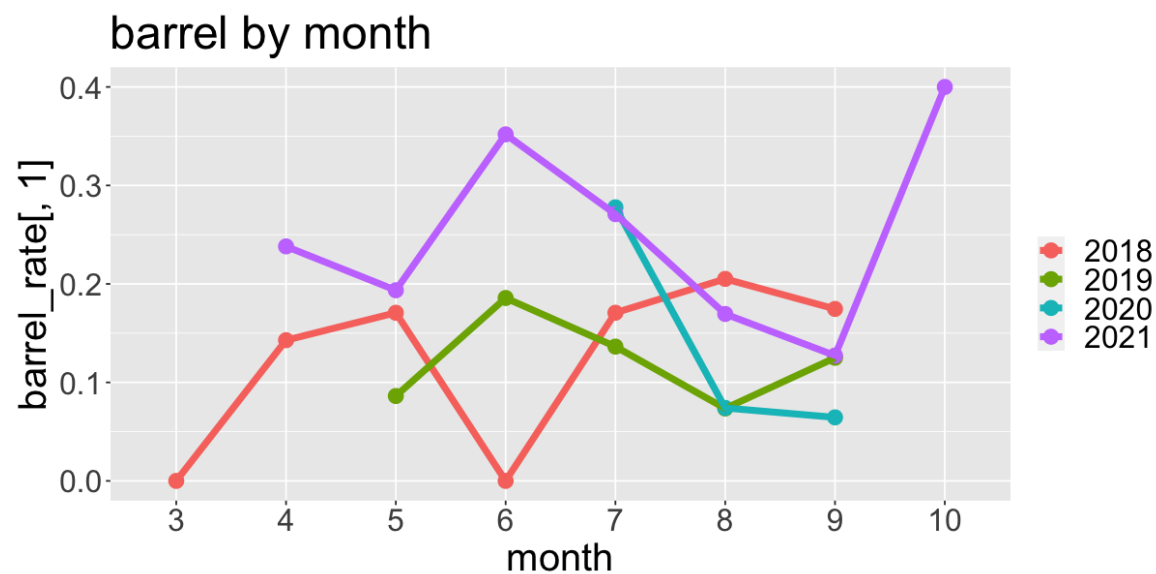
"Removed 4745 rows containing non-finite values (stat_ydensity)."

launch angle by year



In [312]:

```
options(repr.plot.width=10, repr.plot.height=5)
exit_by_type_month %>%
  ggplot(aes(x=factor(month), y=barrel_rate[,1], color=factor(year))) + geom_point(size=4)+geom_line(size=2, aes(group=factor(year)))+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("barrel by month")
```

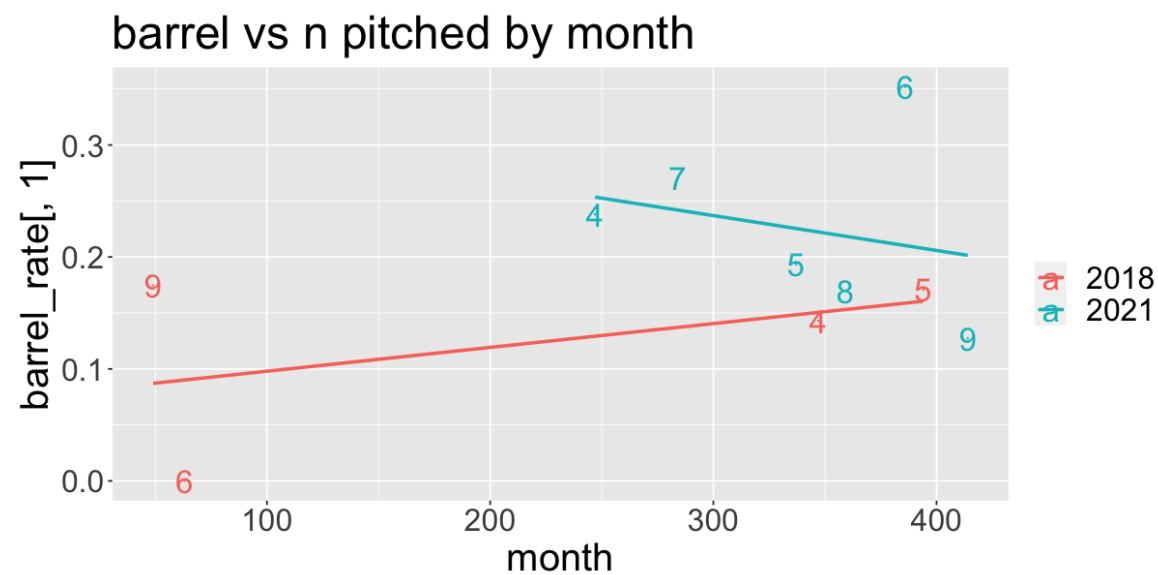


疲労の影響

In [518]:

```
# 月間barrel率と月間投球数
pitch_and_bat %>%
  filter(year!=2020)%>%
  ggplot(aes(x=np, y=barrel_rate[,1], color=factor(year), label=as.character(month))) + geom_point(
    size=0)+stat_smooth(method = "lm", se=F)+
  geom_text(size=7)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("barrel vs n pitched by month")
```

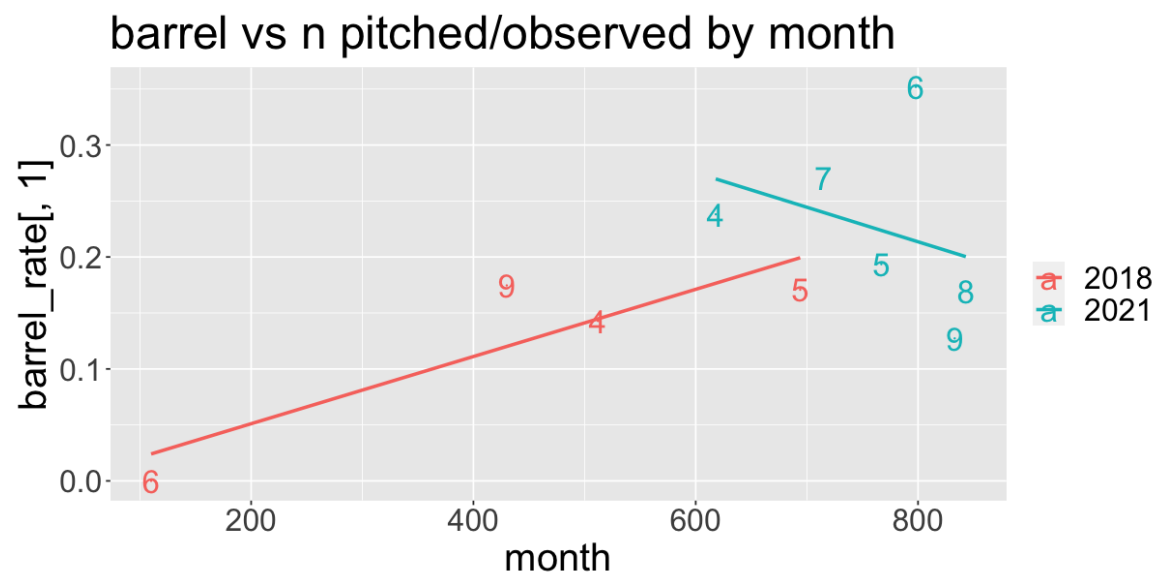
`geom_smooth()` using formula 'y ~ x'



In [519]:

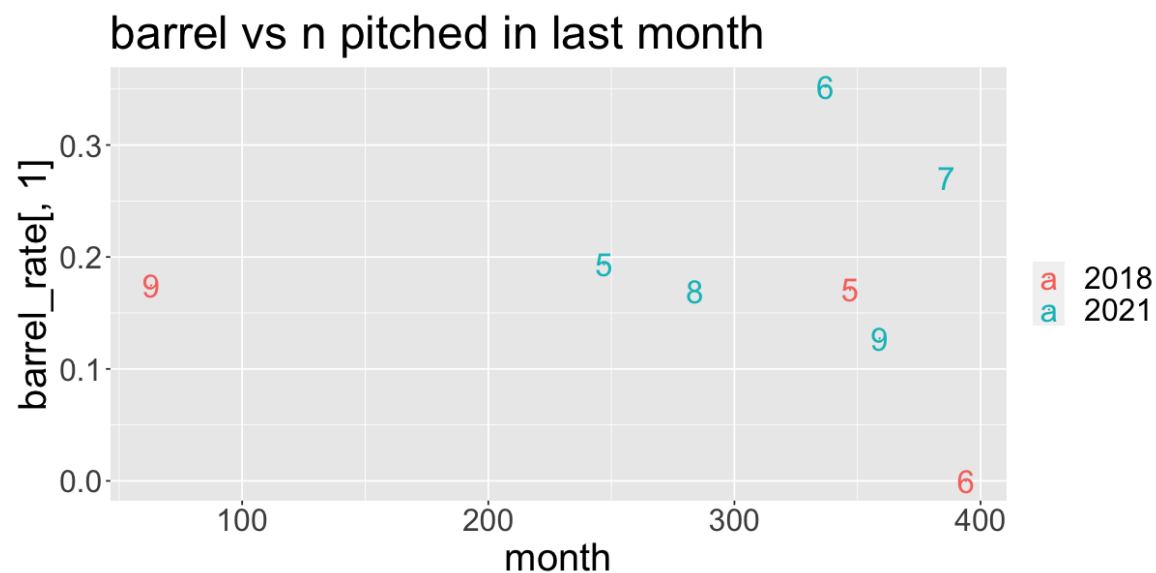
```
# 月間barrel率と月間投球数+打席での球数
pitch_and_bat %>%
  filter(year!=2020)%>%
  ggplot(aes(x=np+npb, y=barrel_rate[,1], color=factor(year), label=as.character(month))) + geom_point(size=0)+stat_smooth(method = "lm", se=F)+
  geom_text(size=7)+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("barrel vs n pitched/observed by month")
```

`geom_smooth()` using formula 'y ~ x'



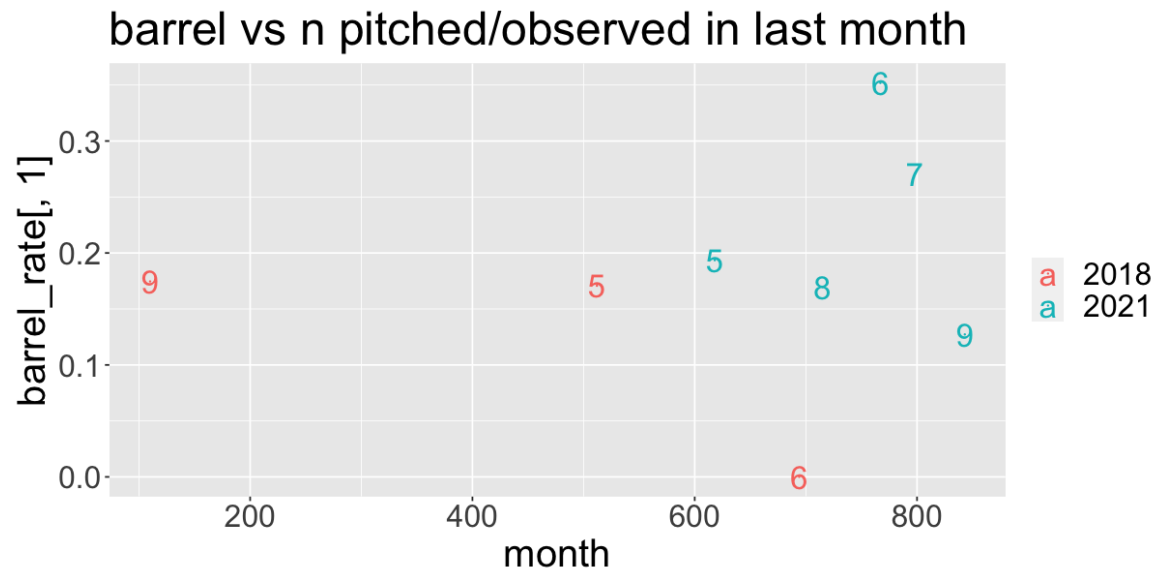
In [527]:

```
# 月間barrel率と前の月の月間投球数
mutate(pitch_and_bat, npd=lag(np), npbd=lag(npb)) %>%
filter(year!=2020 & month!=4)%>%
ggplot(aes(x=npd, y=barrel_rate[,1], color=factor(year), label=as.character(month))) + geom_point
(size=0)+#stat_smooth(method = "lm", se=F)+
geom_text(size=7)+
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("barrel vs n pitched in last month")
```



In [529]:

```
# 月間barrel率と前の月の月間投球数+打席での球数
mutate(pitch_and_bat, npd=lag(np), npbd=lag(npb)) %>%
filter(year!=2020 & month!=4)%>%
ggplot(aes(x=npd+npbd, y=barrel_rate[,1], color=factor(year), label=as.character(month))) + geo
m_point(size=0)+#stat_smooth(method = "lm", se=F)+
geom_text(size=7)+
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("barrel vs n pitched/observed in last month")
```



In [508]:

```
# 6,7月がすごかった！
# 6月の launch angleは低め
```

In [494]:

```
pitch_dates <- as.numeric(rownames(table(all_pitches$datenum)))
pitch_dates_after <- c(pitch_dates + 1, pitch_dates + 2)
#pitch_dates_after
```

In [495]:

```
all_bats$isAfterPitch <- as.numeric(all_bats$datenum %in% pitch_dates_after)
all_bats$isAfterPitch_ <- "after pitch day"
all_bats$isAfterPitch_[all_bats$isAfterPitch==0] <- "the others"
```

In [496]:

```
exit_by_type_month_after <- aggregate(list("barrel_rate"=all_bats$isBarrel,"exit_velo"=all_bats$launch_speed, "launch_angle"=all_bats$launch_angle), list("isAfterPitch"=all_bats$isAfterPitch,"year"=all_bats$year, "month"=all_bats$month), mean_)
exit_by_type_month_after$isAfterPitch_ <- "after pitch day"
exit_by_type_month_after$isAfterPitch_[exit_by_type_month_after$isAfterPitch==0] <- "the others"
```

In [498]:

```
options(repr.plot.width=10, repr.plot.height=8)
all_bats %>%
filter(year==2021)%>%
ggplot(aes(x=factor(month), y=launch_speed, color=factor(isAfterPitch_))) + geom_boxplot()+
geom_point(data = subset(exit_by_type_month_after, exit_by_type_month_after$year==2021),
mapping = aes(x = factor(month), y = exit_velo[,1], color=factor(isAfterPitch_), group = fa
ctor(isAfterPitch_)), size=4)+#, position=position_dodge(width = 0.7))
geom_line(data = subset(exit_by_type_month_after, exit_by_type_month_after$year==2021),
mapping = aes(x = factor(month), y = exit_velo[,1], color=factor(isAfterPitch_), group = fa
ctor(isAfterPitch_)), size=2)+#, position=position_dodge(width = 0.7))
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitl
e("exit velo after pitch day 2018")+
geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

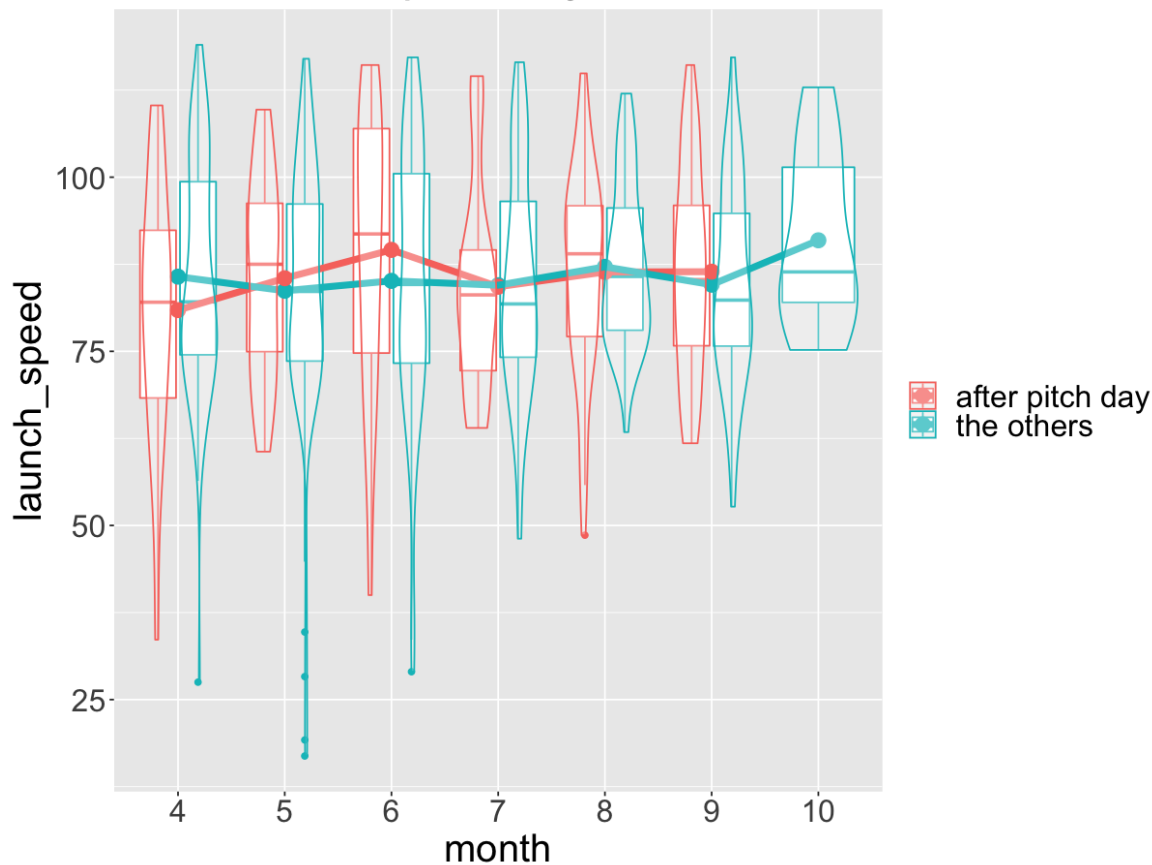
Warning message:

“Removed 1885 rows containing non-finite values (stat_boxplot).”

Warning message:

“Removed 1885 rows containing non-finite values (stat_ydensity).”

exit velo after pitch day 2018



In [499]:

```
options(repr.plot.width=10, repr.plot.height=8)
all_bats %>%
filter(year==2018)%>%
ggplot(aes(x=factor(month), y=launch_speed, color=factor(isAfterPitch_))) + geom_boxplot()+
geom_point(data = subset(exit_by_type_month_after, exit_by_type_month_after$year==2018),
  mapping = aes(x = factor(month), y = exit_velo[,1], color=factor(isAfterPitch_), group = fa
  ctor(isAfterPitch_)), size=4)+#, position=position_dodge(width = 0.7))
geom_line(data = subset(exit_by_type_month_after, exit_by_type_month_after$year==2018),
  mapping = aes(x = factor(month), y = exit_velo[,1], color=factor(isAfterPitch_), group = fa
  ctor(isAfterPitch_)), size=2)+#, position=position_dodge(width = 0.7))
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitl
e("exit velo after pitch day 2018")+
geom_violin(alpha=0.3, position=position_dodge(width=0.8))
```

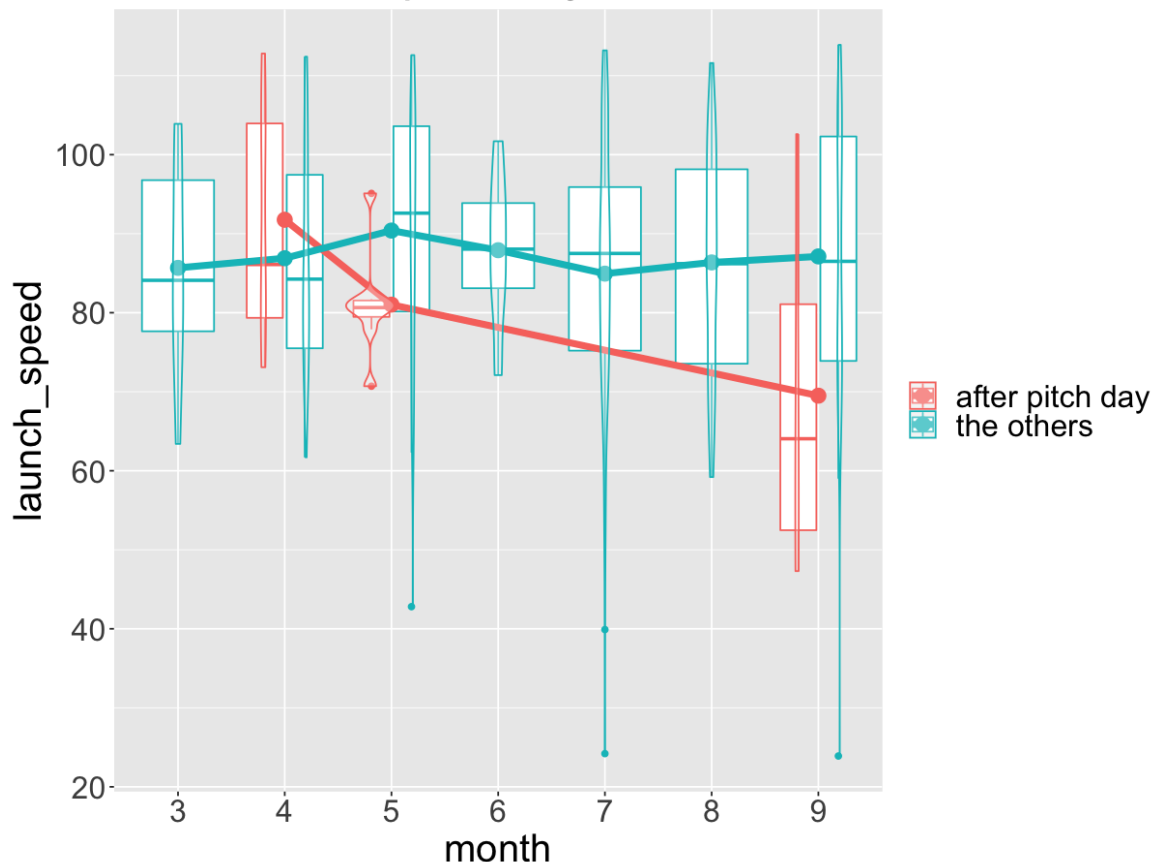
Warning message:

"Removed 1098 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 1098 rows containing non-finite values (stat_ydensity)."

exit velo after pitch day 2018



In [460]:

```
#as.data.frame(table(all_bats$description))
swing <- c("foul", "hit_into_play", "foul_tip", "swinging_strike", "swinging_strike_blocked")
contact <- c("foul", "hit_into_play")

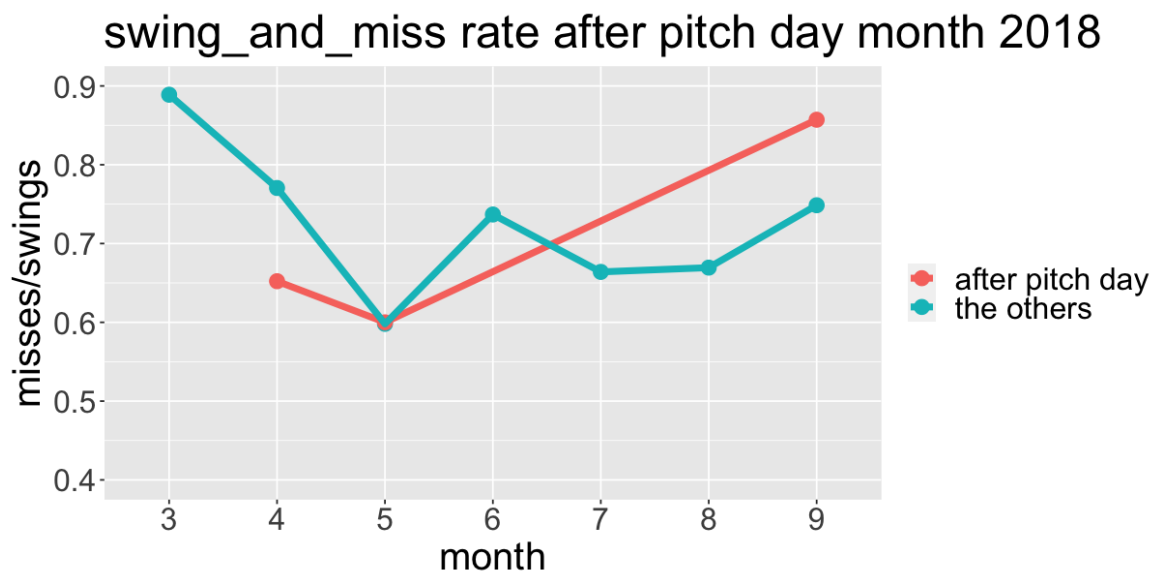
all_bats$isSwing <- as.numeric(all_bats$description %in% swing)
all_bats$isContact <- as.numeric(all_bats$description %in% contact)
```


In [500]:

```
miss_rate_month_after <- aggregate(list("swings"=all_bats$Swing, "misses"=all_bats$Contact), list("isAfterPitch"=all_bats$isAfterPitch,"year"=all_bats$year, "month"=all_bats$month), sum)
miss_rate_month_after$isAfterPitch_ <- "after pitch day"
miss_rate_month_after$isAfterPitch_[miss_rate_month_after$isAfterPitch==0] <- "the others"
```

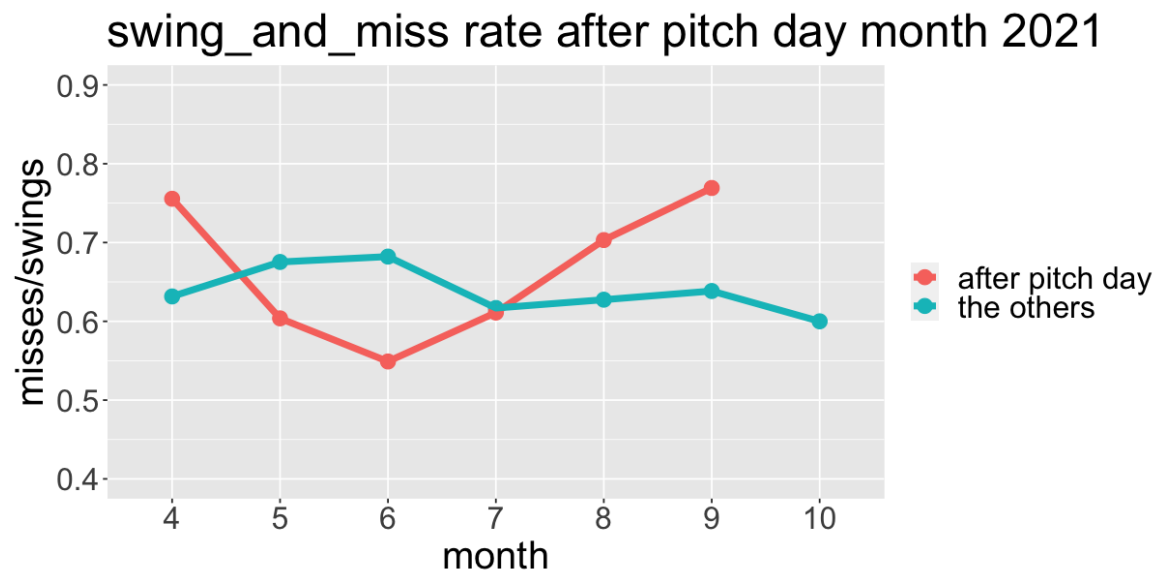
In [531]:

```
options(repr.plot.width=10, repr.plot.height=5)
miss_rate_month_after %>%
  filter(year==2018)%>%
  ggplot(aes(x=factor(month), y=misses/swings, color=factor(isAfterPitch_))) + geom_point(size=4)
+geom_line(size=2, aes(group=factor(isAfterPitch_)))+
  theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("swing_and_miss rate after pitch day month 2018") + ylim(0.4,0.9)
```



In [532]:

```
options(repr.plot.width=10, repr.plot.height=5)
miss_rate_month_after %>%
filter(year==2021)%>%
ggplot(aes(x=factor(month), y=misses/swings, color=factor(isAfterPitch_))) + geom_point(size=4)
+geom_line(size=2, aes(group=factor(isAfterPitch_)))+
theme(text = element_text(size = 24))+xlab("month")+ theme(legend.title=element_blank()) + ggtitle("swing_and_miss rate after pitch day month 2021") + ylim(0.4,0.9)
```



In []:

In []:

In []:

In []: