hw1\_code

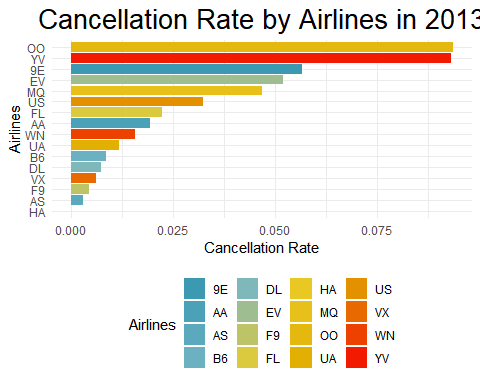
pacman::p\_load(ggplot2, tidyverse, dplyr, plotly, processx, nycflights13)  
setwd("C:/Users/dnskd/Desktop/20Spring/graphics/week1/hw")

airlines\_df <- airlines  
airports\_df <- airports  
flights\_df <- flights  
planes\_df <- planes  
weather\_df <- weather

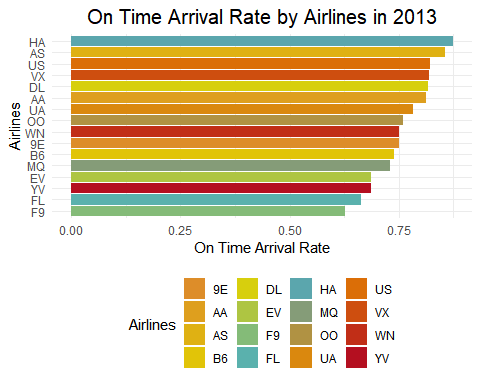
## cancelled 표시  
flights\_df2 <- flights\_df %>% mutate(cancelled = ifelse(is.na(dep\_time) == TRUE, 1, 0))  
flights\_df3 <- flights\_df2 %>% left\_join(airlines\_df, by = "carrier")  
  
## 결항율  
## palette 써서 color에 customization을 주고 싶다.  
library(wesanderson)

## Warning: package 'wesanderson' was built under R version 3.6.3

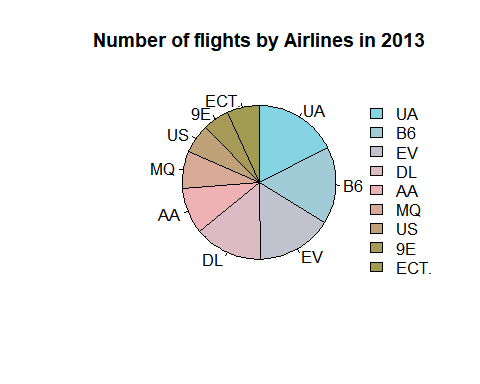
pal <- wes\_palette("Zissou1", 16, type = c("continuous"))  
  
flights\_df3 %>% group\_by(carrier) %>% summarise(cancelled\_rate = mean(cancelled)) %>%  
 ggplot(aes(x = reorder(carrier, cancelled\_rate), y = cancelled\_rate, fill = carrier)) + geom\_bar(stat = "identity") + theme\_bw() + theme\_minimal()+  
 labs(title = "Cancellation Rate by Airlines in 2013", x = "Airlines", y = "Cancellation Rate") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "bottom") +  
 coord\_flip() + scale\_fill\_manual(name = "Airlines", values = pal)



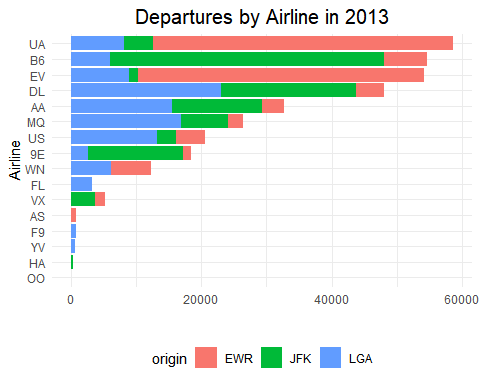
pal2 <- wes\_palette("FantasticFox1", 16, type = c("continuous"))  
flights\_df4 <- flights\_df3 %>% mutate(on.time= ifelse(arr\_delay < 16, 1, 0))  
  
flights\_df4 %>% group\_by(carrier) %>% summarise(on\_time\_rate = mean(on.time, na.rm = T)) %>%  
 ggplot(aes(x = reorder(carrier, on\_time\_rate), y = on\_time\_rate, fill = carrier)) + geom\_bar(stat = "identity") + theme\_bw() + theme\_minimal()+  
 labs(title = "On Time Arrival Rate by Airlines in 2013", x = "Airlines", y = "On Time Arrival Rate") +  
 theme(plot.title = element\_text(hjust = 0.5, size = 15), legend.position = "bottom") + coord\_flip()+  
 scale\_fill\_manual(name = "Airlines", values = pal2)



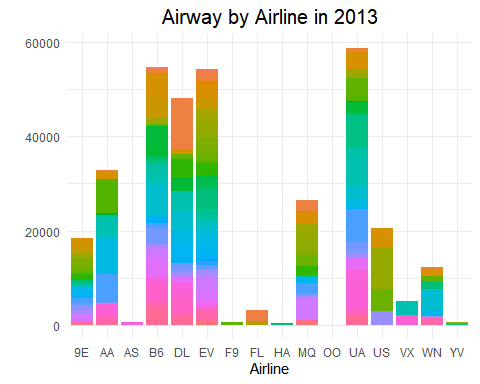
pal3 <- wes\_palette("Moonrise3", 16, type = c("continuous"))  
count\_flights <- flights\_df %>% group\_by(carrier) %>% summarise(total = n()) %>% arrange(-total)  
top8 <- count\_flights[1:8, ]  
worst <- count\_flights[9:16, ] %>% summarise(total = sum(total))  
worst$carrier <- 'ECT.'  
worst <- worst %>% select(carrier, total)  
count <- rbind(top8, worst)  
pie(count$total, count$carrier, col = pal3, clockwise = TRUE, main = "Number of flights by Airlines in 2013")  
legend("bottomright", legend = count$carrier, bty = "n", fill = pal3)



flights\_df %>% group\_by(carrier, origin) %>% summarise(total = n()) %>%  
 ggplot(aes(x = reorder(carrier, total) , y = total, fill = origin)) + geom\_bar(position = "stack", stat = "identity") + theme\_bw() + theme\_minimal()+  
 labs(title = "Departures by Airline in 2013", x = "Airline", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 15), legend.position = "bottom") + coord\_flip()



flights\_df %>% group\_by(carrier, dest) %>% summarise(total = n()) %>%  
 ggplot(aes(carrier, total, fill = dest)) + geom\_bar(position = "stack", stat = "identity") + theme\_bw() + theme\_minimal()+  
 labs(title = "Airway by Airline in 2013", x = "Airline", y = " ") + theme(legend.position = "none", plot.title = element\_text(hjust = 0.5, size = 15))



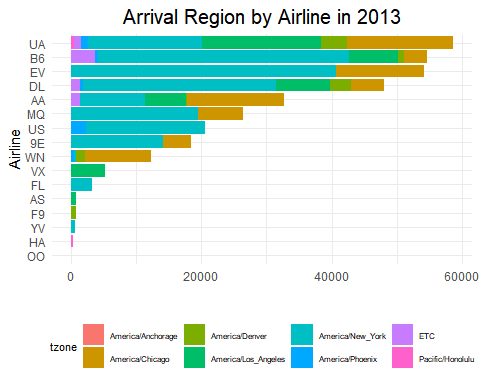
airports\_df %>% filter(is.na(tzone))

## # A tibble: 3 x 8  
## faa name lat lon alt tz dst tzone  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr> <chr>  
## 1 EEN Dillant Hopkins Airport 72.3 42.9 149 -5 A <NA>   
## 2 LRO Mount Pleasant Regional-Faiso~ 32.5 -79.5 12 -5 A <NA>   
## 3 YAK Yakutat 59.3 -139. 33 -9 A <NA>

airports\_df[airports\_df$faa %in% c("EEN", "LRO"), "tzone"] <- "America/New\_York"  
airports\_df[airports\_df$faa == "YAK", "tzone"] <- "America/Yakutat"  
airports\_df %>% filter(faa %in% c("EEN", "LRO", "YAK"))

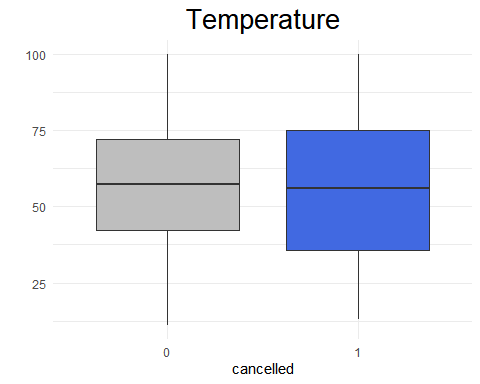
## # A tibble: 3 x 8  
## faa name lat lon alt tz dst tzone   
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr> <chr>   
## 1 EEN Dillant Hopkins Airport 72.3 42.9 149 -5 A America/Ne~  
## 2 LRO Mount Pleasant Regional~ 32.5 -79.5 12 -5 A America/Ne~  
## 3 YAK Yakutat 59.3 -139. 33 -9 A America/Ya~

flights\_airport <- flights\_df %>% left\_join(airports\_df, by = c("dest" = "faa"))  
flights\_airport[is.na(flights\_airport$tzone), "tzone"] <- "ETC"  
  
df1 <- flights\_airport %>% group\_by(carrier, tzone) %>% summarise(total = n())   
df2 <- flights\_airport %>% group\_by(carrier) %>% summarise(ttotal = n())  
df1 %>% left\_join(df2 , by = "carrier") %>% ggplot(aes(x = reorder(carrier, ttotal), y = total, fill = tzone)) + geom\_bar(position = "stack", stat = "identity") + theme\_bw() + theme\_minimal()+  
 labs(title = "Arrival Region by Airline in 2013", x = "Airline", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 15), legend.position = "bottom", legend.title = element\_text(size = 8), legend.text = element\_text(size = 6 ) ) + coord\_flip()



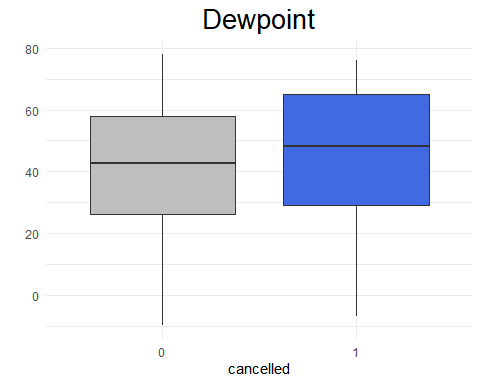
flights\_weather <- flights\_df %>% left\_join(weather\_df, by = c("year", "month", "day", "hour","origin", "time\_hour"))  
flights\_weather <- flights\_weather %>% mutate(cancelled = ifelse(is.na(dep\_time) == TRUE, 1, 0))  
flights\_weather %>% ggplot(aes(as.factor(cancelled), temp)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Temperature", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none") + scale\_fill\_manual(values = c("gray", "royalblue"))

## Warning: Removed 1573 rows containing non-finite values (stat\_boxplot).



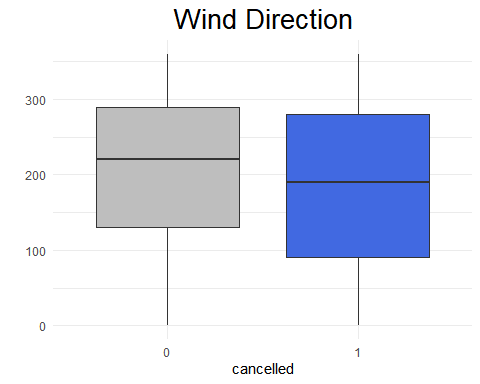
flights\_weather %>% ggplot(aes(as.factor(cancelled), dewp)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Dewpoint", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none") + scale\_fill\_manual(values = c("gray", "royalblue")) # 결항때 dewp가 좀더 높다.

## Warning: Removed 1573 rows containing non-finite values (stat\_boxplot).



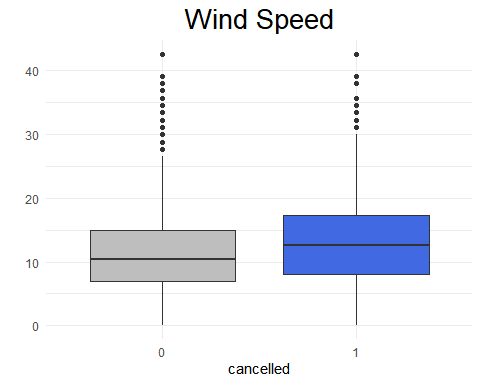
flights\_weather %>% ggplot(aes(as.factor(cancelled), wind\_dir)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Wind Direction", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none") + scale\_fill\_manual(values = c("gray", "royalblue"))

## Warning: Removed 9796 rows containing non-finite values (stat\_boxplot).



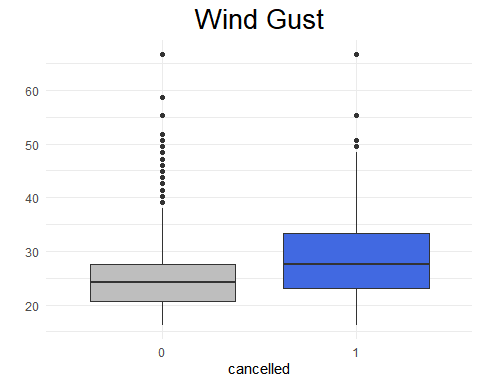
flights\_weather %>% ggplot(aes(as.factor(cancelled), wind\_speed)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Wind Speed", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none") + scale\_fill\_manual(values = c("gray", "royalblue"))

## Warning: Removed 1634 rows containing non-finite values (stat\_boxplot).



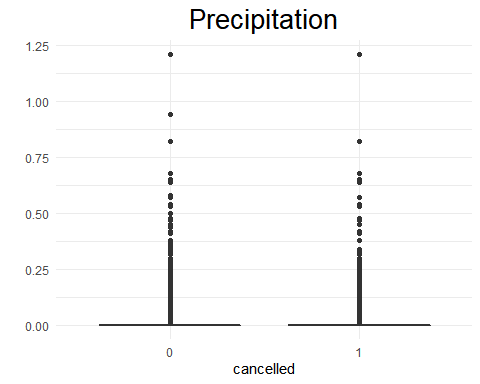
flights\_weather %>% ggplot(aes(as.factor(cancelled), wind\_gust)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Wind Gust", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none") + scale\_fill\_manual(values = c("gray", "royalblue"))

## Warning: Removed 256391 rows containing non-finite values (stat\_boxplot).



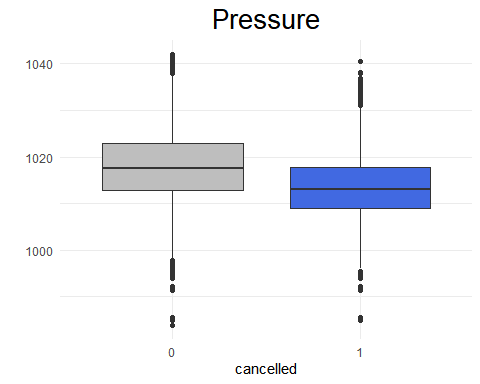
flights\_weather %>% ggplot(aes(as.factor(cancelled), precip)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Precipitation", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none")

## Warning: Removed 1556 rows containing non-finite values (stat\_boxplot).



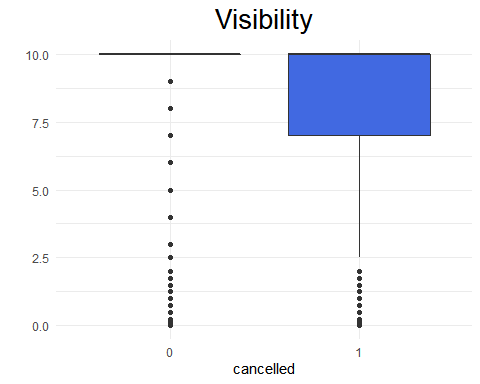
flights\_weather %>% ggplot(aes(as.factor(cancelled), pressure)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Pressure", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none") + scale\_fill\_manual(values = c("gray", "royalblue"))

## Warning: Removed 38788 rows containing non-finite values (stat\_boxplot).

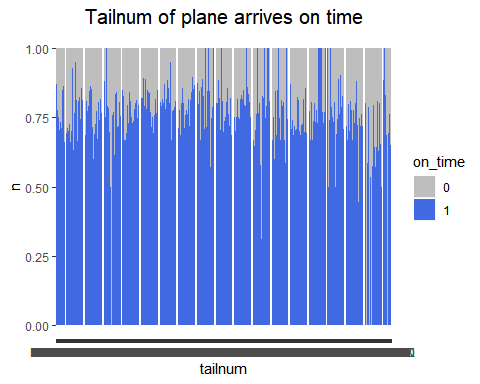


flights\_weather %>% ggplot(aes(as.factor(cancelled), visib)) + geom\_boxplot(aes(group= as.factor(cancelled), fill = as.factor(cancelled))) + theme\_bw() + theme\_minimal() + labs(title = "Visibility", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20)) + labs(title = "Visibility", x = "cancelled", y = " ") + theme(plot.title = element\_text(hjust = 0.5, size = 20), legend.position = "none") + scale\_fill\_manual(values = c("gray", "royalblue"))

## Warning: Removed 1556 rows containing non-finite values (stat\_boxplot).

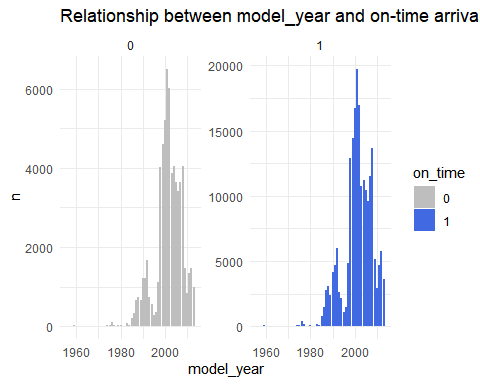


flights\_planes <- flights\_df %>% left\_join(planes\_df, by = "tailnum") %>% rename(model\_year = year.y) %>%  
 mutate(on.time = ifelse(arr\_delay < 16, 1, 0))   
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(tailnum, on.time) %>% summarise(n = n()) %>%  
 ggplot(aes(tailnum, n)) + geom\_bar(stat = "identity", aes(fill = as.factor(on.time)), position = "fill") +  
 scale\_fill\_manual("on\_time", values = c("gray", "royalblue")) + labs(title = "Tailnum of plane arrives on time") + theme(plot.title = element\_text(hjust = 0.5, size = 15))

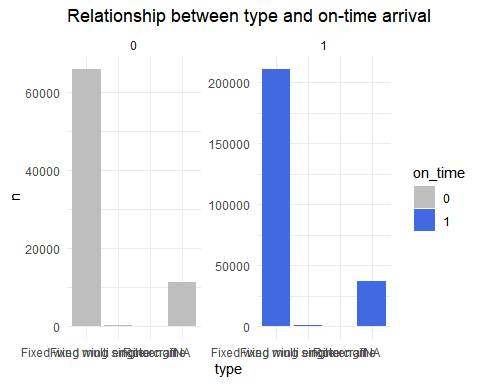


## year  
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(model\_year, on.time) %>% summarise(total = n()) %>%  
 ggplot(aes(model\_year, total)) + geom\_bar(stat = "identity", aes(fill = as.factor(on.time))) +  
 scale\_fill\_manual("on\_time", values = c("gray", "royalblue")) + labs(title = "Relationship between model\_year and on-time arrival", x = "model\_year", y = "n")+ theme(plot.title = element\_text(hjust = 0.5, size = 15), axis.text.x = element\_text(angle = 45, hjust = 1)) + theme\_minimal()+  
 facet\_wrap(~on.time, scale = "free\_y")

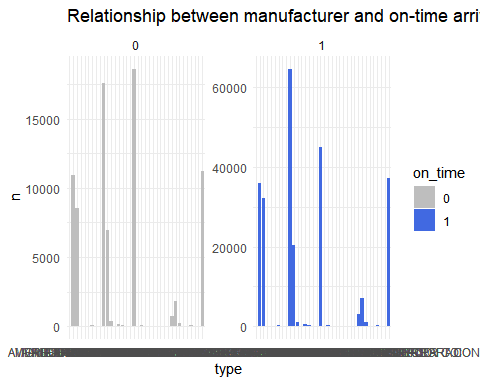
## Warning: Removed 2 rows containing missing values (position\_stack).



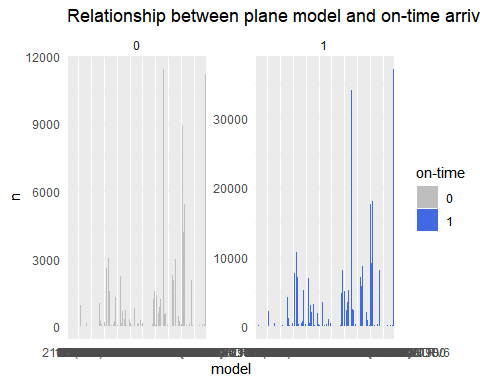
## type  
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(type, on.time) %>% summarise(total = n()) %>%  
 ggplot(aes(type, total)) + geom\_bar(stat = "identity", aes(fill = as.factor(on.time))) +  
 scale\_fill\_manual("on\_time", values = c("gray", "royalblue")) + labs(title = "Relationship between type and on-time arrival", x = "type", y = "n")+ theme(plot.title = element\_text(hjust = 0.5, size = 15), axis.text.x = element\_text(angle = 45, hjust = 1)) + theme\_minimal()+  
 facet\_wrap(~on.time, scale = "free\_y")



## manufacturer  
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(manufacturer, on.time) %>% summarise(total = n()) %>%  
 ggplot(aes(manufacturer, total)) + geom\_bar(stat = "identity", aes(fill = as.factor(on.time))) +  
 scale\_fill\_manual("on\_time", values = c("gray", "royalblue")) + labs(title = "Relationship between manufacturer and on-time arrival", x = "type", y = "n")+ theme(plot.title = element\_text(hjust = 0.5, size = 15), axis.text.x = element\_text(angle = 45, hjust = 1)) + theme\_minimal()+  
 facet\_wrap(~on.time, scale = "free\_y")

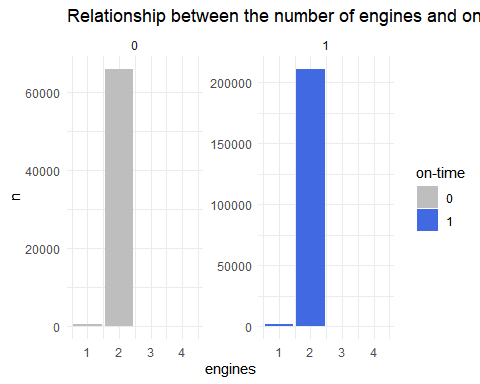


## model  
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(model, on.time) %>% summarise(total = n()) %>%  
 ggplot(aes(model, total)) + geom\_bar(stat = "identity", aes(fill = as.factor(on.time))) + facet\_wrap(~on.time, scale = "free\_y")+  
 scale\_fill\_manual("on-time", values = c("gray", "royalblue")) + labs(title = "Relationship between plane model and on-time arrival", x = "model", y = "n")+  
 theme(plot.title = element\_text(hjust = 0.5, size = 15)) + theme\_minimal()



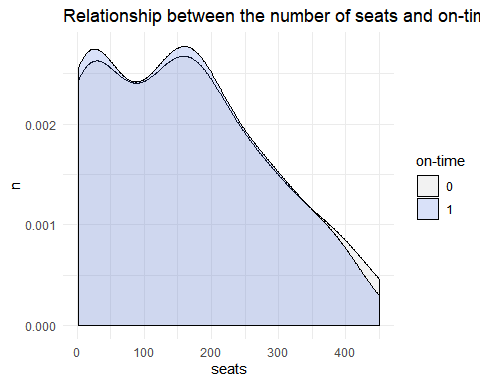
## engines  
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(engines, on.time) %>% summarise(total = n()) %>%  
 ggplot(aes(engines, total))+geom\_bar(stat = "identity", aes(fill = as.factor(on.time))) + facet\_wrap(~on.time, scale = "free\_y")+  
 scale\_fill\_manual("on-time", values = c("gray", "royalblue")) + labs(title = "Relationship between the number of engines and on-time arrival", x = "engines", y = "n")+  
 theme(plot.title = element\_text(hjust = 0.5, size = 10)) + theme\_minimal()

## Warning: Removed 2 rows containing missing values (position\_stack).



## seats  
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(seats, on.time) %>% summarise(total = n()) %>%  
 ggplot(aes(seats, total)) + geom\_density(aes(y = ..density.., fill = as.factor(on.time)), alpha = 0.2)+  
 scale\_fill\_manual("on-time", values = c("gray", "royalblue")) + labs(title = "Relationship between the number of seats and on-time arrival", x = "seats", y = "n")+  
 theme(plot.title = element\_text(hjust = 0.5, size = 10)) + theme\_minimal()

## Warning: Removed 2 rows containing non-finite values (stat\_density).



## engines  
flights\_planes %>% filter(!is.na(on.time)) %>% group\_by(engine, on.time) %>% summarise(total = n()) %>%  
 ggplot(aes(engine, total))+geom\_bar(stat = "identity", aes(fill = as.factor(on.time))) + facet\_wrap(~on.time, scale = "free\_y")+  
 scale\_fill\_manual("on-time", values = c("gray", "royalblue")) + labs(title = "Relationship between type of engine and on-time arrival", x = "engines", y = "n")+  
 theme(plot.title = element\_text(hjust = 0.5, size = 10)) + theme\_minimal()

