Analysis_big_cars_epa

Serdar Korur

10/21/2019

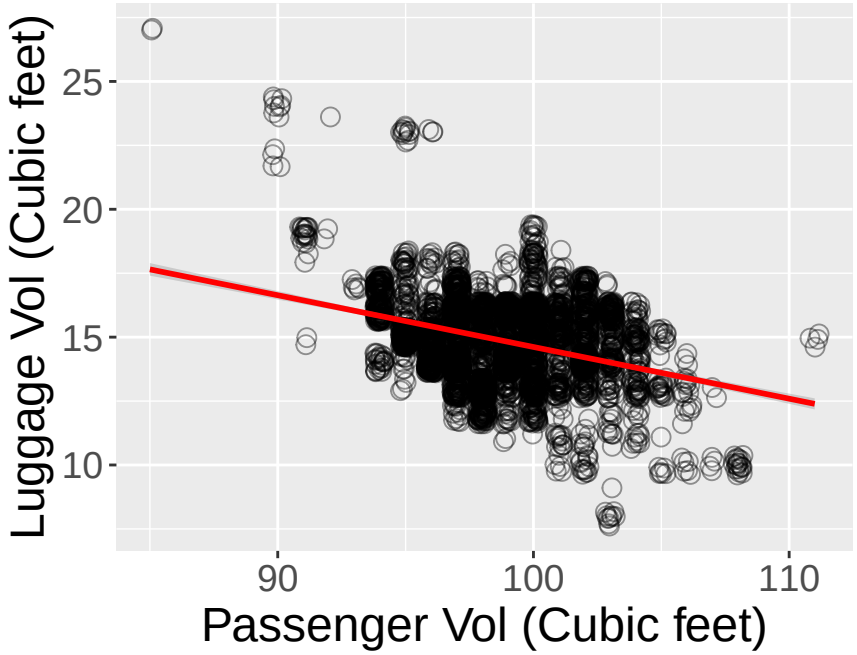
```
library(tidyverse) # ggplot2, dplyr, tidyr, readr,
                   # purrr, tibble, stringr, forcats
big_epa_cars <- read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2019
dim(big_epa_cars)
```

```
## [1] 41804    83
```

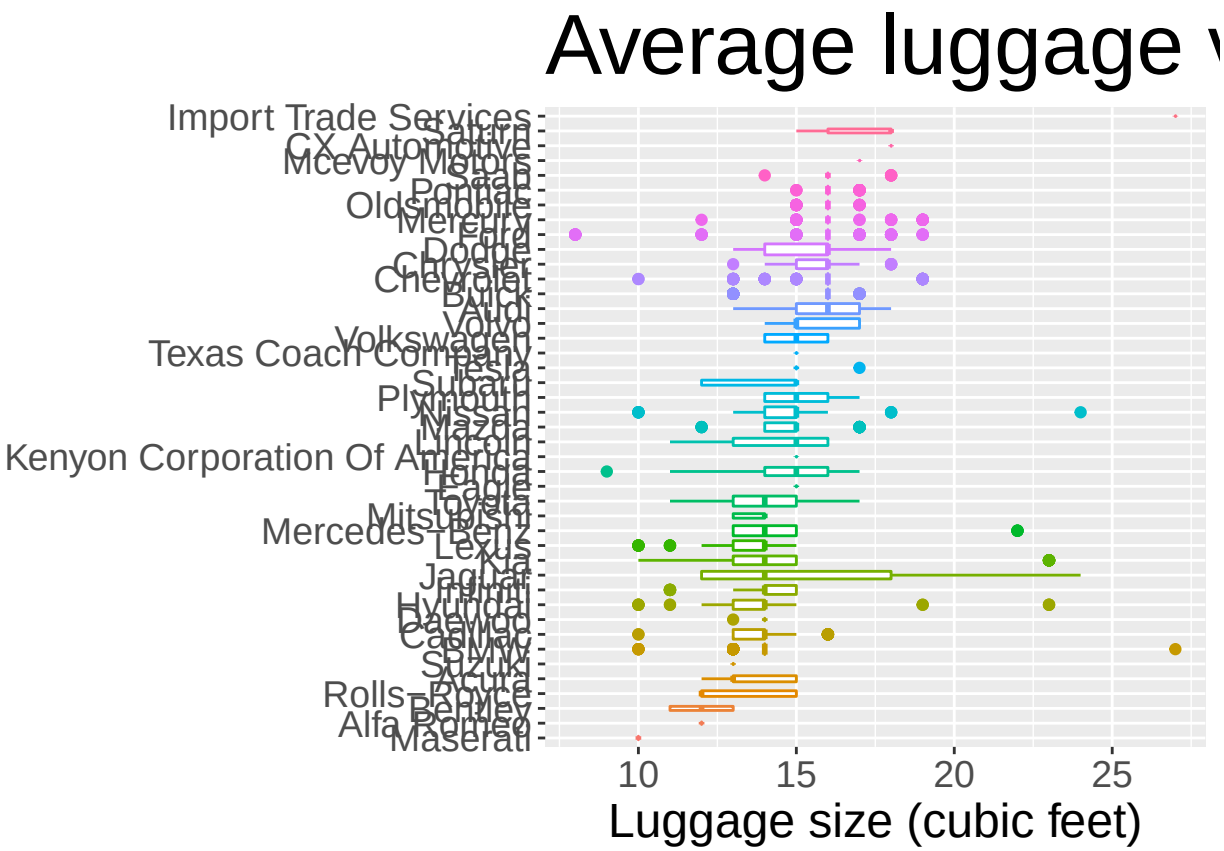
```
big_sub <- big_epa_cars %>%
  select(fuelType, year, make, model, VClass, hlv, hpv,lv4,pv4,displ)
```

```
posn.j <- position_jitter(width=0.2)
big_sw <- big_sub %>%
  filter(VClass == "Midsize Cars" & pv4 > 75 & lv4 > 6)

big_sw %>%
  ggplot(aes(x=pv4, y=lv4)) +
  geom_point(shape=21, alpha=0.4,size =3, position = posn.j) +
  theme(plot.caption=element_text(size=11), text = element_text(size=18),    plot.title = element_text(s
  geom_smooth(method = "lm", color ="red") +
  coord_fixed() +
    labs(x = "Passenger Vol (Cubic feet)", y = "Luggage Vol (Cubic feet)", title = "Luggage space negati
```

Luggage space negativel correlates with passenge



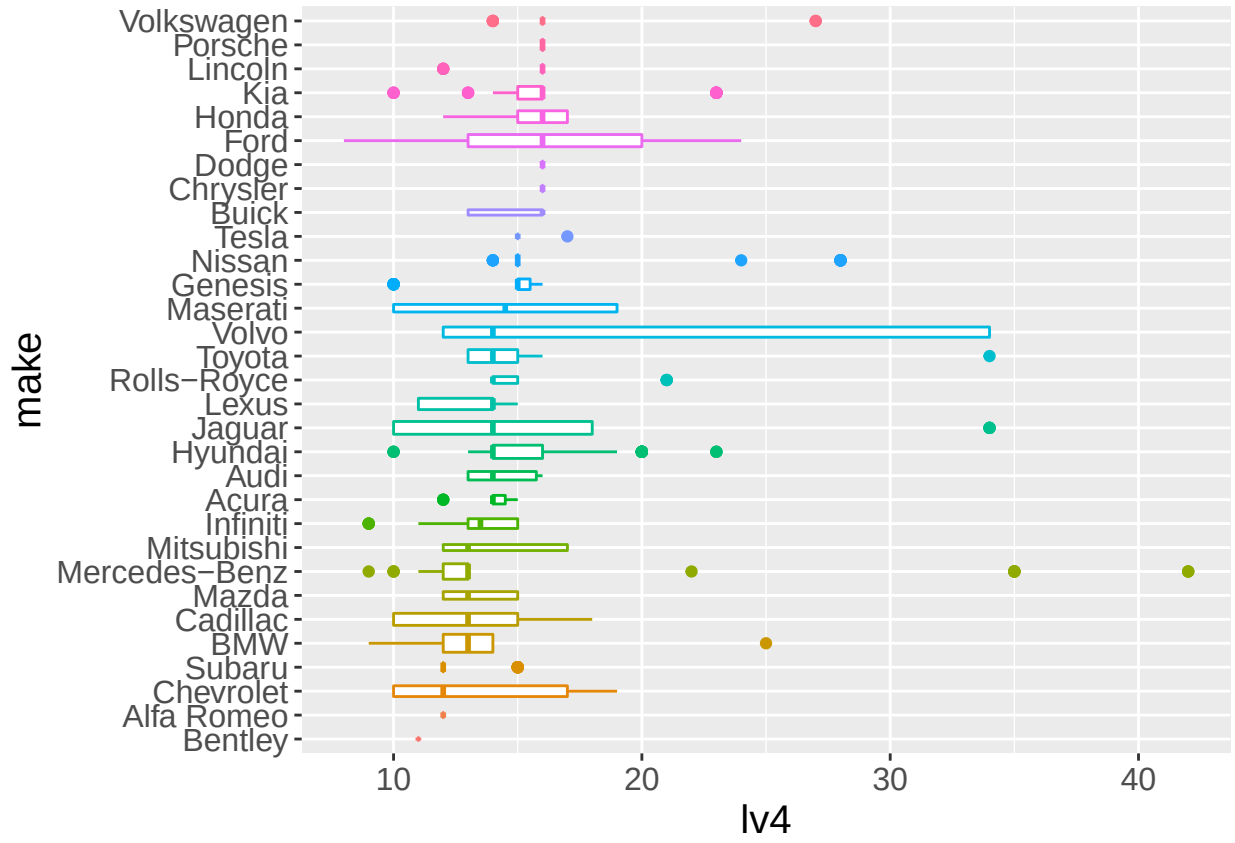

```
pp <- big_sw %>%
  mutate(make = fct_reorder(make, lv4)) %>%
  ggplot(aes(x=make, y=lv4, col=make)) +
  geom_boxplot(varwidth=TRUE) +
  theme(plot.caption=element_text(size=11), text = element_text(size=18),    plot.title = element_text(
  coord_flip() +
  labs(x = element_blank(), y = "Luggage size (cubic feet)", title = "Average luggage volumes in Midsiz
  pp
```



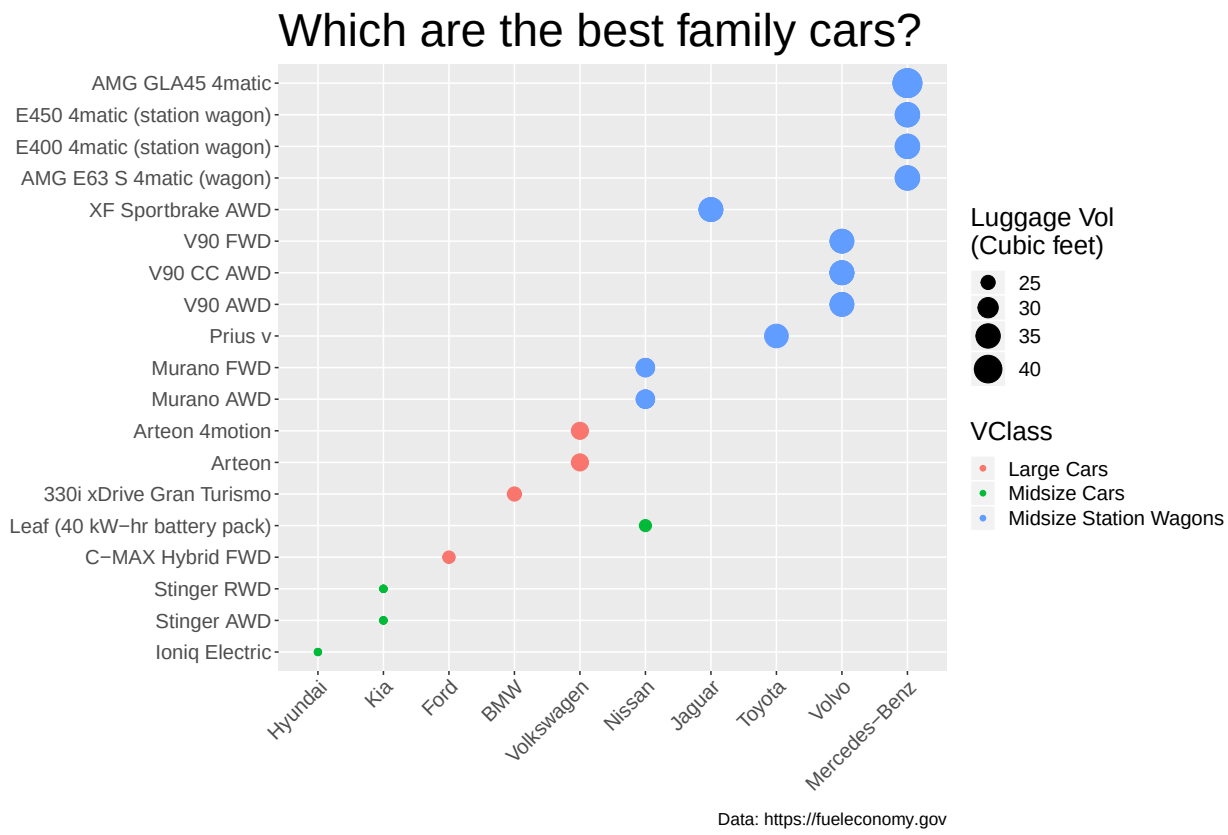

```
big_filtered <- big_sub %>%
  filter(VClass %in% c("Large Cars", "Compact Cars", "Midsize Cars",
                       "Midsize Station Wagons", "Midsize-Large Station Wagons",
                       "Minivan - 2WD", "Minivan - 4WD")) %>%
  group_by(make) %>%
  mutate(n=n()) %>%
  filter(n > 10) %>%
  ungroup()

dim(big_filtered)
```

```
## [1] 14710    11
```

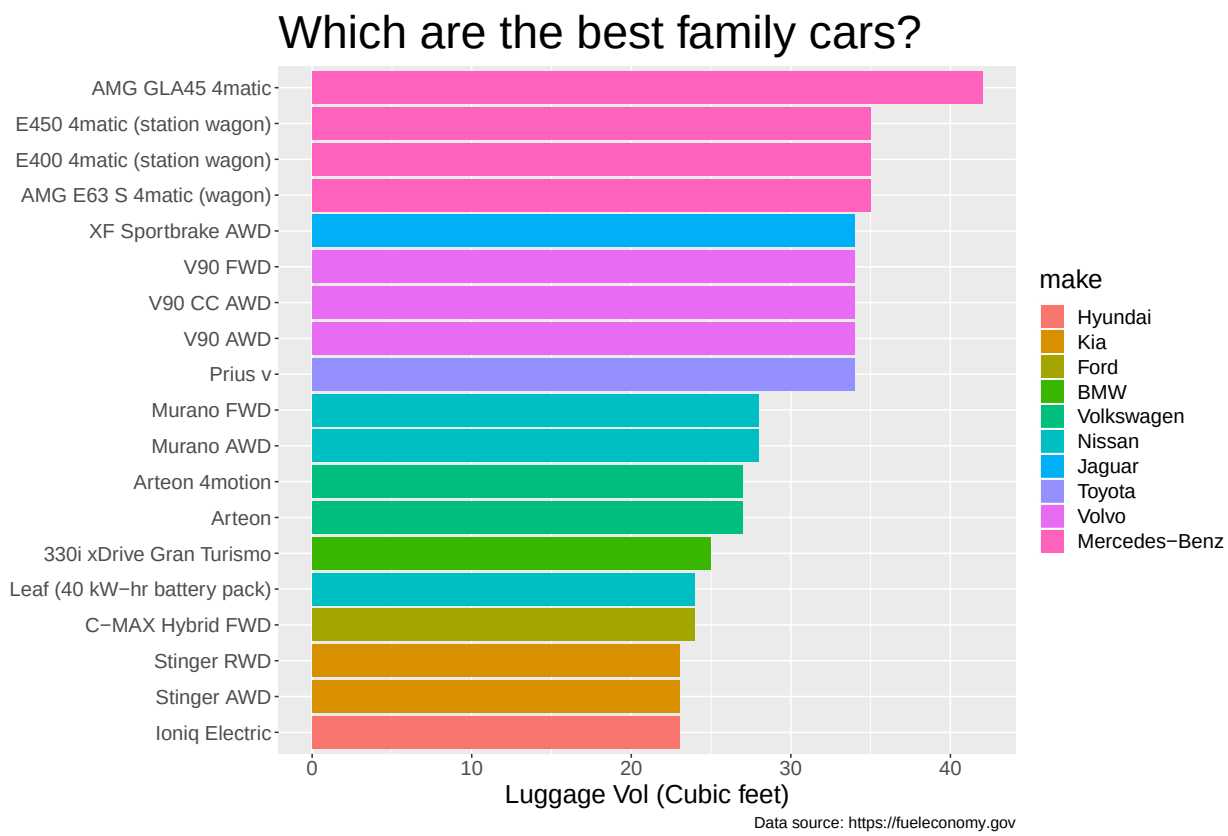
```
# Cars ordered with luggage volume, but not older than 5 years
# and lv4 bigger than 5

q <- big_filtered %>%
  filter(year > 2016, lv4 > 5) %>%
  mutate(make = fct_reorder(make, lv4)) %>%
  ggplot(aes(x=make, y=lv4, col=make)) +
  geom_boxplot(varwidth=TRUE) +
  theme(text = element_text(size=15), legend.position = "none") +
  coord_flip()
q
```



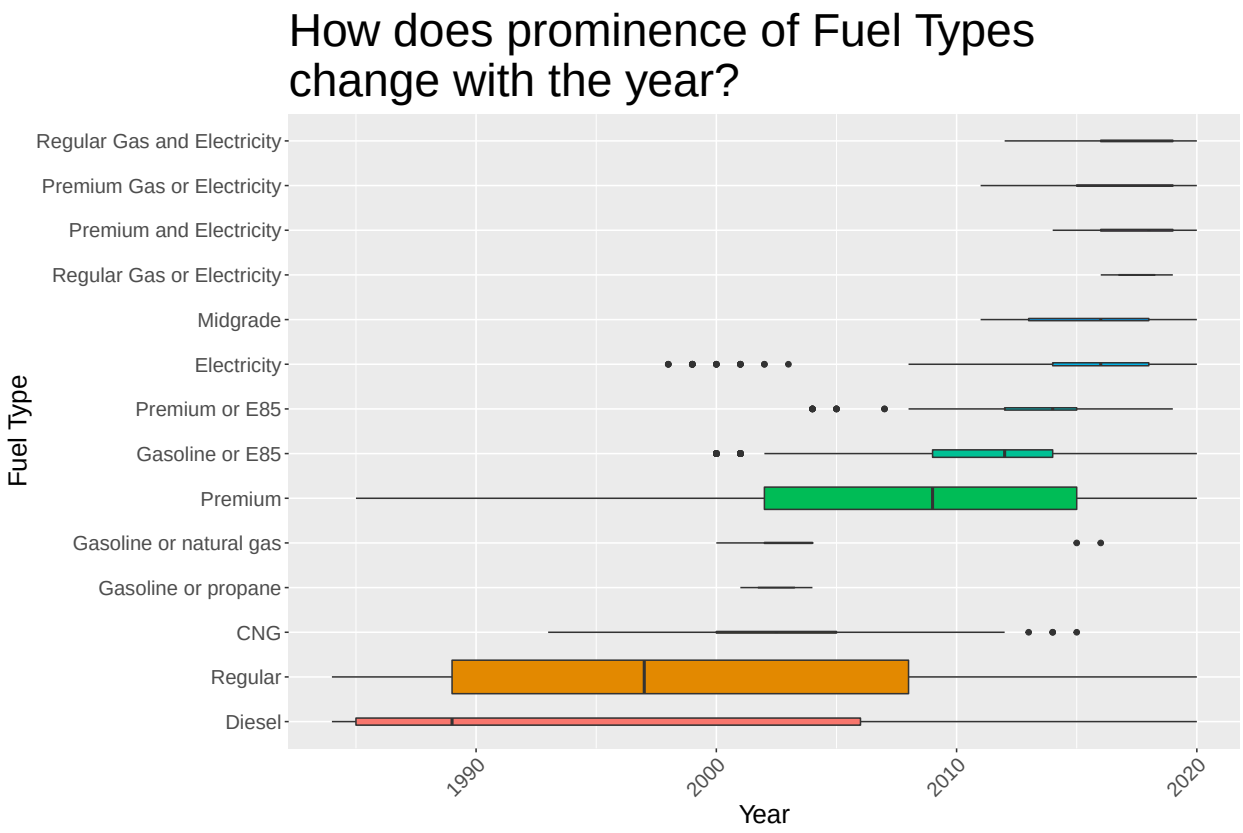

```
boot_space <- big_filtered %>%
  filter(year > 2016) %>%
  arrange(desc(lv4)) %>%
  top_n(50, lv4)

# Top family cars - geom_point()
bs <- boot_space %>%
  mutate(model = fct_reorder(model, lv4)) %>%
  mutate(make = fct_reorder(make, lv4)) %>%
  ggplot(aes(x=make,y= model, size=lv4, col=VClass)) +
  geom_point() +
  theme(plot.caption=element_text(size=12),axis.text.x=element_text(angle=45, hjust=1),text = element_t
labs(caption= "Data: https://fueleconomy.gov", size="Luggage Vol\n(Cubic feet)", x = element_blank(), y
bs
```



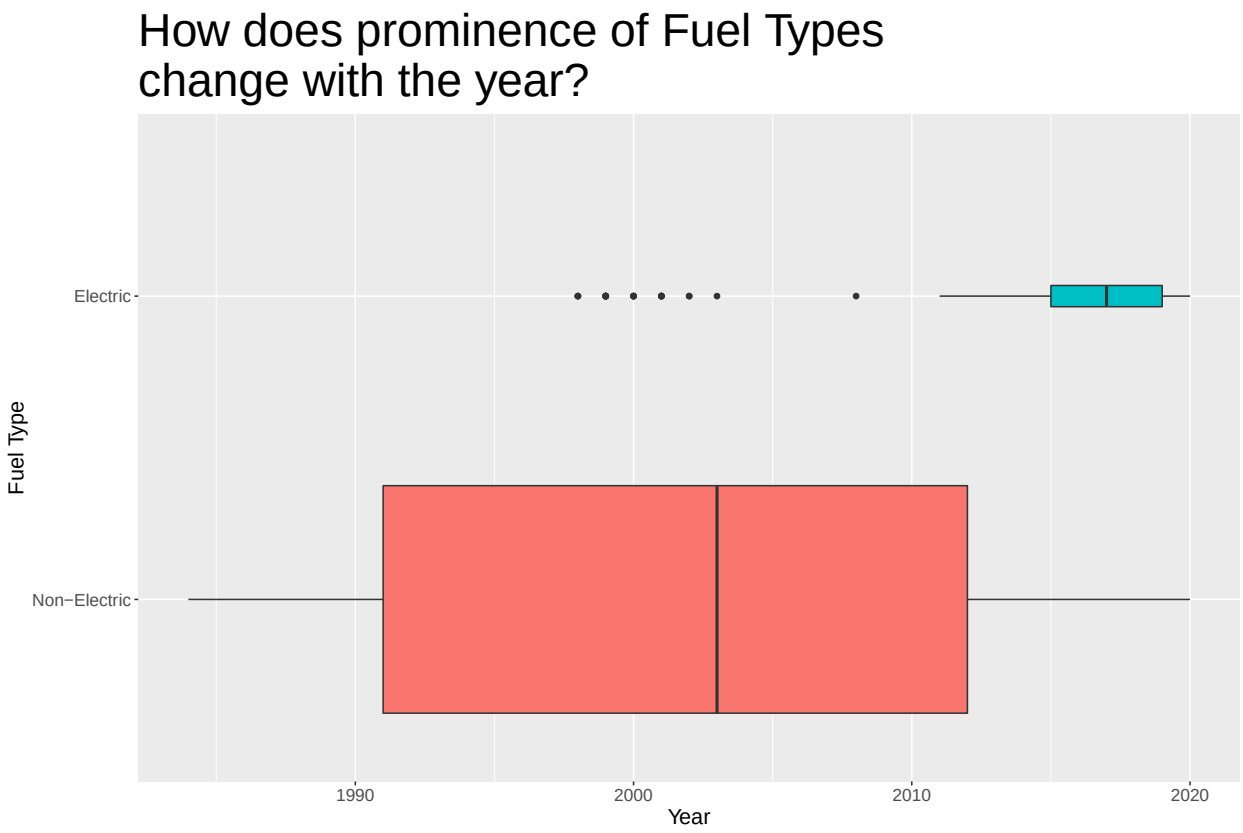

```
# Top family cars - geom_Col()
bs_col <- boot_space %>%
    mutate(model = fct_reorder(model, lv4)) %>%
    mutate(make = fct_reorder(make, lv4)) %>%
    ggplot(aes(x=model, y=lv4, fill=make)) +
    geom_col(position="dodge")+coord_flip() +
    theme(plot.caption=element_text(size=11), text = element_text(size=18), plot.title = element_text(si
labs(caption= "Data source: https://fueleconomy.gov", size="Luggage Vol\n(Cubic feet)", x = element_blan
    scale_size(range=c(2, 9))
bs_col
```




```
# Using Varwidth: Ordered
pp  <- big_epa_cars %>%
  mutate(fuelType=fct_reorder(fuelType, year)) %>%
  ggplot(aes(x=fuelType, y =year, fill=fuelType)) +
  geom_boxplot(varwidth=TRUE) + theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip() +
  theme(legend.position = "none", text = element_text(size=18),
        plot.title = element_text(size=32)) +
  labs(x = "Fuel Type", y = "Year", title = "How does prominence of Fuel Types \nchange with the year?"
pp
```




```
# Grouped: Electric vs no electric:
big_epa_cars$fuelType <- ifelse( big_epa_cars$fuelType %in% c("Regular Gas and Electricity", "Premium Ga

pp  <- big_epa_cars %>%
  mutate(fuelType=fct_reorder(fuelType, year)) %>%
  ggplot(aes(x=fuelType, y =year, fill=fuelType)) +
  geom_boxplot(varwidth=TRUE) +
  coord_flip() +
  theme(text = element_text(size=18),
        plot.title = element_text(size=32), legend.position = "none")+
  theme(text = element_text(size=15)) +
  labs(x = "Fuel Type", y = "Year", title = "How does prominence of Fuel Types \nchange with the year?"
pp
```




```
big3 <- big_epa_cars %>% group_by(year, fuelType) %>% mutate(n = n())

big3 %>%
    ggplot(aes(x=n, y =year, col=fuelType)) +
    geom_point() +
    theme(legend.position = c(0.9,0.9), legend.title= element_blank(), legend.background = element_blank
    theme(plot.title = element_text(size=32), text = element_text(size=15)) + coord_flip() +
    labs(x = "Number of Car models", y = "Year", title = "How does the Numbers of Electric vs Non Electr
```

How does the Numbers of Electric vs Non Electric cars change by year?