12.1 CS_1 (NYC Flights)

Soru 1

```
# A tibble: 1 × 3
  year.y manufacturer model
   <int> <chr>        <chr>
1   2002 BOEING       767-424ER
> |
```

Soru 2

```
Correlation computed with
• Method: 'pearson'
• Missing treated using: 'pairwise.complete.obs'
# A tibble: 4 × 5
  term       air_time    temp wind_speed   humid
  <chr>         <dbl>   <dbl>      <dbl>   <dbl>
1 air_time     NA      -0.0367     0.0263  0.0405
2 temp         -0.0367 NA         -0.140   0.0374
3 wind_speed    0.0263 -0.140      NA      -0.187
4 humid         0.0405  0.0374    -0.187   NA
> |
```

Bu korelasyondan bazı çıkarımlar yapabiliriz. Örneğin, sıcaklık (temp) ile uçuş süresi (air_time) arasında negatif bir korelasyon olduğu için, sıcaklık arttığında uçuş süresinin daha kısa olacağını, sıcaklık azaldığında ise uçuş süresinin daha uzun olacağını söyleyebiliriz. Ancak, korelasyonun zayıf olduğunu göz önünde bulundurursak, bu ilişkiyi kesin olarak söylemek mümkün değildir.

Soru 3

```
> data3<- right_join(airlines,flights,by="carrier")
> data3 %>% filter(dep_delay>0) %>% na.omit() %>% grou
# A tibble: 16 × 3
   name                       mean median
   <chr>                     <dbl>  <dbl>
 1 AirTran Airways Corporation  40.6   16
 2 Alaska Airlines Inc.         31.5   12
 3 American Airlines Inc.       37.2   16
 4 Delta Air Lines Inc.         37.3   16
 5 Endeavor Air Inc.            48.5   26
 6 Envoy Air                    44.7   27
 7 ExpressJet Airlines Inc.     50.2   31
 8 Frontier Airlines Inc.       45.2   18
 9 Hawaiian Airlines Inc.       44.8    5
10 JetBlue Airways              39.7   20
11 Mesa Airlines Inc.           52.9   29.5
12 SkyWest Airlines Inc.        58     40
13 Southwest Airlines Co.       34.8   15
14 US Airways Inc.              32.9   16
15 United Air Lines Inc.        29.8   12
16 Virgin America               34.2   10
> |
```

Soru 4

```
> flights %>% filter(dep_delay
# A tibble: 1 x 1
   mean
  <dbl>
1  39.4
>
```

12.2 CS_2 (Sean 'Lahman' Baseball Database)

Soru 1

```
  nameFirst nameLast
1     Barry    Bonds
>
```

Soru 2

```
[1] 9719
>
```

Soru 3

```
> data2<-Salaries %>% left_join(t,by="playerID")
> which.max(data2$salary)==which.max(data2$num.of.aw)
[1] FALSE
>
> slice(data2,which.max(data2$salary))     #max salary
  yearID teamID lgID  playerID   salary num.of.aw
1   2009    NYA   AL rodrial01 33000000        31
>
> slice(data2,which.max(data2$num.of.aw))
  yearID teamID lgID  playerID salary num.of.aw
1   1986    PIT   NL bondsba01  60000        47
>
```

Soru 4

```
                             Var1 Freq
1                     TSN All-Star 1525
2           Baseball Magazine All-Star 1520
3                       Gold Glove 1204
4                     Silver Slugger  792
5                Most Valuable Player  208
6                  Rookie of the Year  154
7              TSN Pitcher of the Year  151
8                     Cy Young Award  126
9           Reliever of the Year Award   94
10              TSN Player of the Year   92
11 TSN Major League Player of the Year   89
12             TSN Fireman of the Year   88
13                    Babe Ruth Award   78
14                   World Series MVP   71
15          Lou Gehrig Memorial Award   69
16                  All-Star Game MVP   62
17                       Hutch Award   55
18              Roberto Clemente Award   54
19                    Hank Aaron Award   50
20                             NLCS MVP   49
21                             ALCS MVP   43
22               Pitching Triple Crown   39
23         Comeback Player of the Year   36
24            TSN Reliever of the Year   36
25                        TSN Guide MVP   33
26                     Platinum Glove   26
27                  Branch Rickey Award   23
28                          Triple Crown   17
29                 Outstanding DH Award    8
30                       SIlver Slugger    5
> |
```

12.3 CS_3 (Diamonds)

Soru 1

```
# A tibble: 1 × 1
  depth
  <dbl>
1  62.6
> |
```

Soru 2

```
# A tibble: 1 × 12
  carat cut   color clarity depth table price     x     y     z discount new.price
  <dbl> <ord> <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>    <dbl>     <dbl>
1  0.23 Ideal E     SI2      61.5    55   326  3.95  3.98  2.43     9.78      316.
> |
```

Soru 3

```
# A tibble: 1 x 12
  carat cut   color clarity depth table price     x     y     z discount new.price
  <dbl> <ord> <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>    <dbl>     <dbl>
1  0.23 Ideal E     SI2      61.5    55   326  3.95  3.98  2.43     9.78      316.
> 
```

Soru 4

```
> diamonds %>% group_by(cut) %>% summarise(mean.pr=mean(price))
# A tibble: 5 x 2
  cut       mean.pr
  <ord>       <dbl>
1 Fair        4359.
2 Good        3929.
3 Very Good   3982.
4 Premium     4584.
5 Ideal       3458.
> 
> # or
> 
> diamonds %>% filter(cut=="Ideal") %>% summarise(mean(price))
# A tibble: 1 x 1
  `mean(price)`
          <dbl>
1         3458.
> 
```

Soru 5

```
> diamonds %>% group_by(cut) %>% summarise(mean.pr=mean(price)) %>%  filter(mean.pr==min(mean.pr)) %>% select(cut,mean.pr)
# A tibble: 1 x 2
  cut   mean.pr
  <ord>   <dbl>
1 Ideal   3458.
> 
```