markdown_practice

2023-01-15

Load libraries

Load in tidyverse and janitor libraries

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'janitor'
##
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

Read csv file

Load a csv into a dataframe [Titanic dataset]

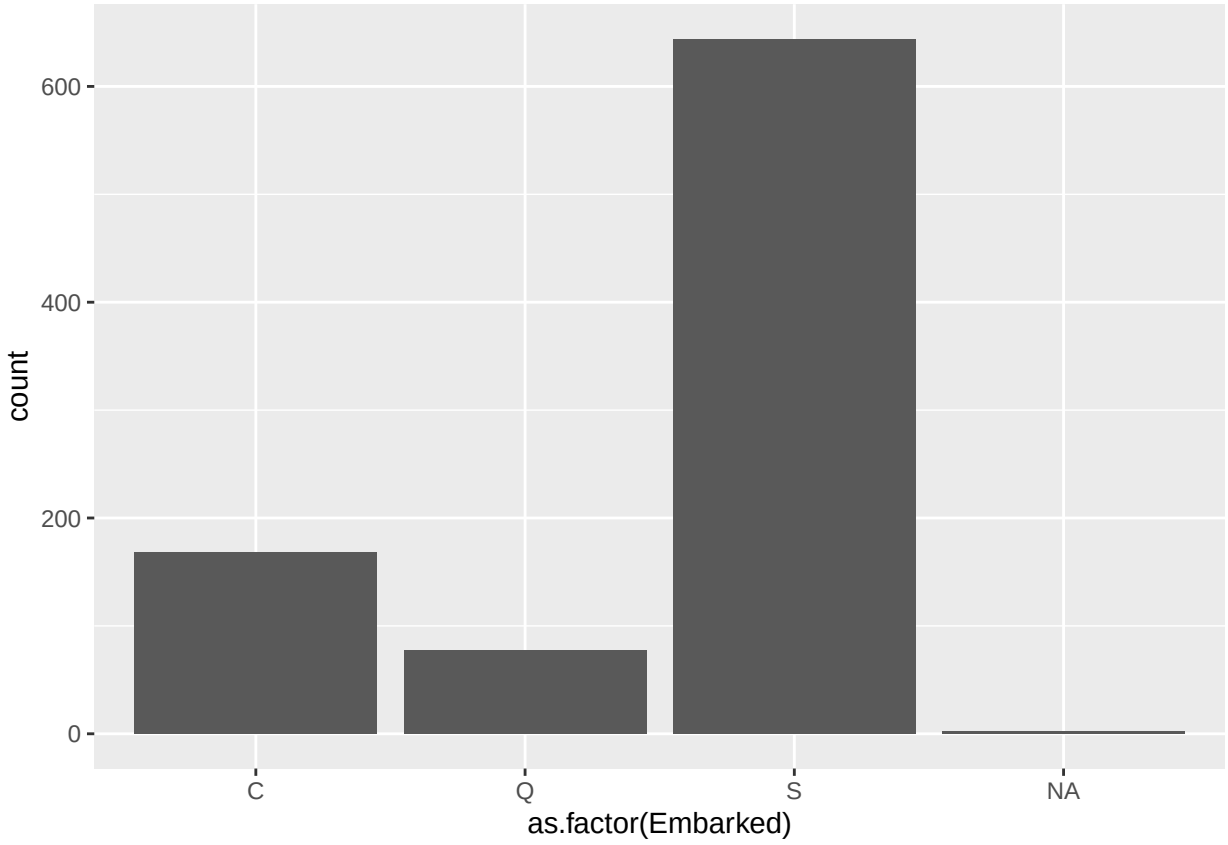
```
titanic_file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/Discussion/R_Basics/titanic.csv'
# read.csv method
titanic = read.csv(titanic_file_path, na.strings='')

# view the dataframe
glimpse(titanic)
```

```
## Rows: 891
## Columns: 12
## $ PassengerId <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,~
## $ Survived    <int> 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1~
## $ Pclass      <int> 3, 1, 3, 1, 3, 3, 1, 3, 3, 2, 3, 1, 3, 3, 3, 2, 3, 2, 3, 3~
## $ Name        <chr> "Braund, Mr. Owen Harris", "Cumings, Mrs. John Bradley (Fl~
## $ Sex         <chr> "male", "female", "female", "female", "male", "male", "mal~
## $ Age         <dbl> 22, 38, 26, 35, 35, NA, 54, 2, 27, 14, 4, 58, 20, 39, 14, ~
## $ SibSp       <int> 1, 1, 0, 1, 0, 0, 0, 3, 0, 1, 1, 0, 0, 1, 0, 0, 4, 0, 1, 0~
## $ Parch       <int> 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 1, 0, 0, 5, 0, 0, 1, 0, 0, 0~
## $ Ticket      <chr> "A/5 21171", "PC 17599", "STON/O2. 3101282", "113803", "37~
## $ Fare        <dbl> 7.2500, 71.2833, 7.9250, 53.1000, 8.0500, 8.4583, 51.8625,~
## $ Cabin       <chr> NA, "C85", NA, "C123", NA, NA, "E46", NA, NA, NA, "G6", "C~
## $ Embarked    <chr> "S", "C", "S", "S", "S", "Q", "S", "S", "S", "C", "S", "S"~
```

Data Visualization

Bar Chart

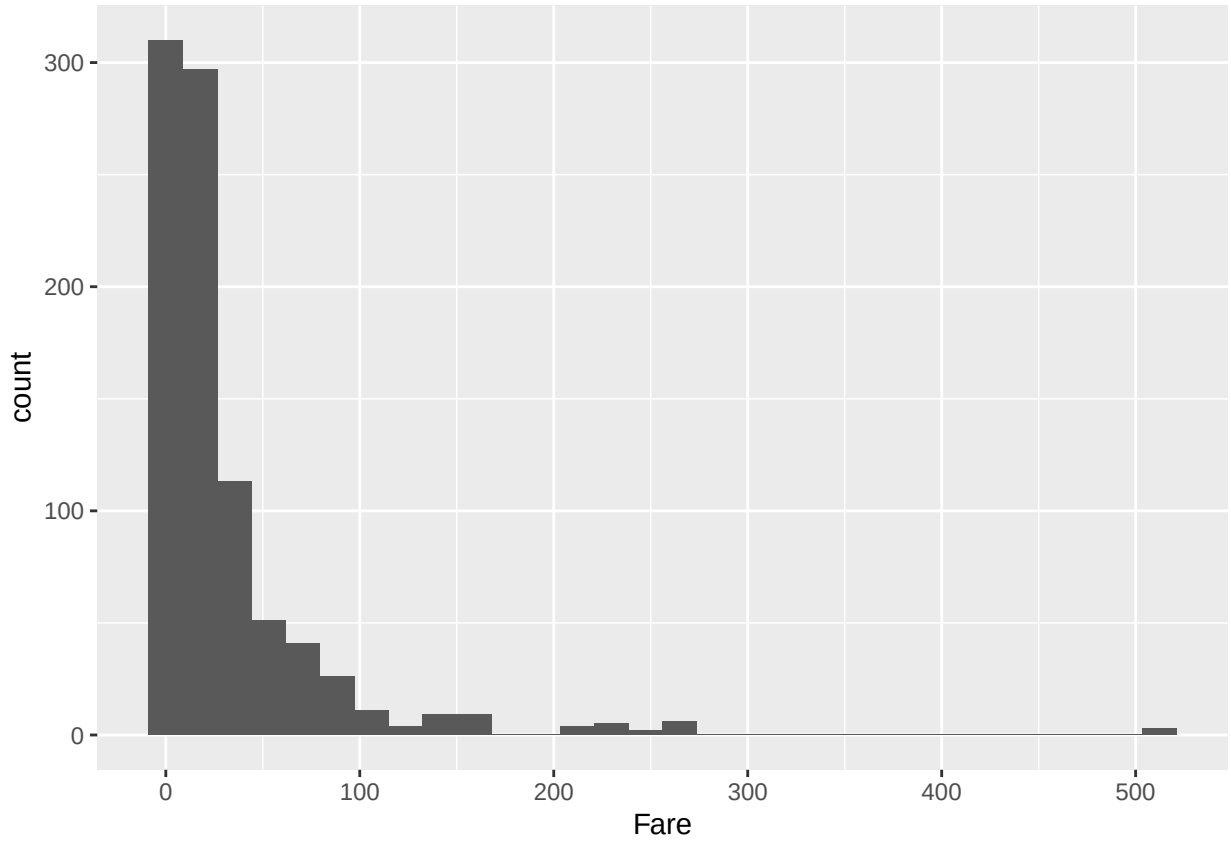
```
# Bargraph of each of the Embarked classes
ggplot(
  data=titanic,
  aes(x=as.factor(Embarked)) # aes is short for aesthetics
) +
geom_bar(na.rm=TRUE)
```



Historgram

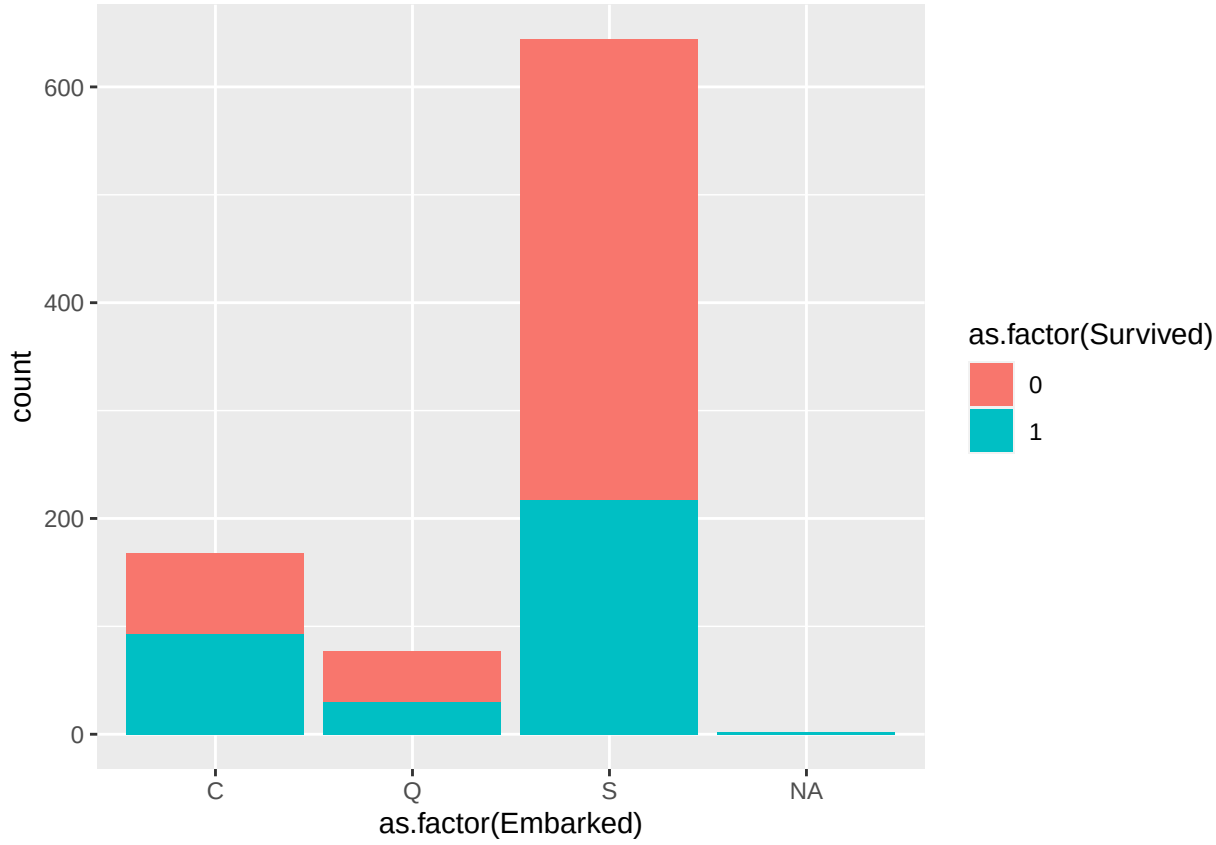
```
# Histogram of the fare paid
ggplot(
  data=titanic,
  aes(x=Fare)
) +
geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Bargraph

```
# Bargraph of each of the Embarked classes with survived amount
ggplot(
  data=titanic,
  aes(x=as.factor(Embarked),
      fill=as.factor(Survived))
) + geom_bar()
```



Dataframe data transformations [Penguins dataset]

```
penguins_file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/Discussion/R_Basics/penguins_full
penguins = read.csv(penguins_file_path, na.strings='')
glimpse(penguins)
```

```
## Rows: 344
## Columns: 10
## $ X                    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15~
## $ species              <chr> "Adelie", "Adelie", "Adelie", "Adelie", "Adelie",~
## $ island               <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen~
## $ year                 <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2~
## $ bill_length_mm       <chr> "39.1", "39.5", "40.3", "NA", "36.7", "39.3", "38~
## $ bill_depth_mm        <chr> "18.7", "17.4", "18", "NA", "19.3", "20.6", "17.8~
## $ flipper_length_mm    <chr> "181", "186", "195", "NA", "193", "190", "181", "~
## $ body_mass_g          <chr> "3750", "3800", "3250", "NA", "3450", "3650", "36~
## $ above_average_weight <chr> "0", "0", "0", "NA", "0", "0", "0", "1", "0", "1"~
## $ sex                  <chr> "male", "female", "female", "NA", "female", "male~
```

chr -> numeric data conversion

bill_length_mm and **bill_depth_mm** need to be converted to numeric data types

```
penguins$bill_length_mm=as.numeric(penguins$bill_length_mm)
```

```
## Warning: NAs introduced by coercion
```

```
penguins$bill_depth_mm = as.numeric(penguins$bill_depth_mm)
```

```
## Warning: NAs introduced by coercion
```

```
glimpse(penguins)
```

```
## Rows: 344
## Columns: 10
## $ X                    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15~
## $ species              <chr> "Adelie", "Adelie", "Adelie", "Adelie", "Adelie",~
## $ island               <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen~
## $ year                 <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2~
## $ bill_length_mm       <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.~
## $ bill_depth_mm        <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.~
## $ flipper_length_mm    <chr> "181", "186", "195", "NA", "193", "190", "181", "~
## $ body_mass_g          <chr> "3750", "3800", "3250", "NA", "3450", "3650", "36~
## $ above_average_weight <chr> "0", "0", "0", "NA", "0", "0", "0", "1", "0", "1"~
## $ sex                  <chr> "male", "female", "female", "NA", "female", "male~
```

chr (categorical) -> factor (numerical representation of class value)

species a class variable needs to be converted to a factor variable

```
penguins$species=as.factor(penguins$species)
glimpse(penguins)
```

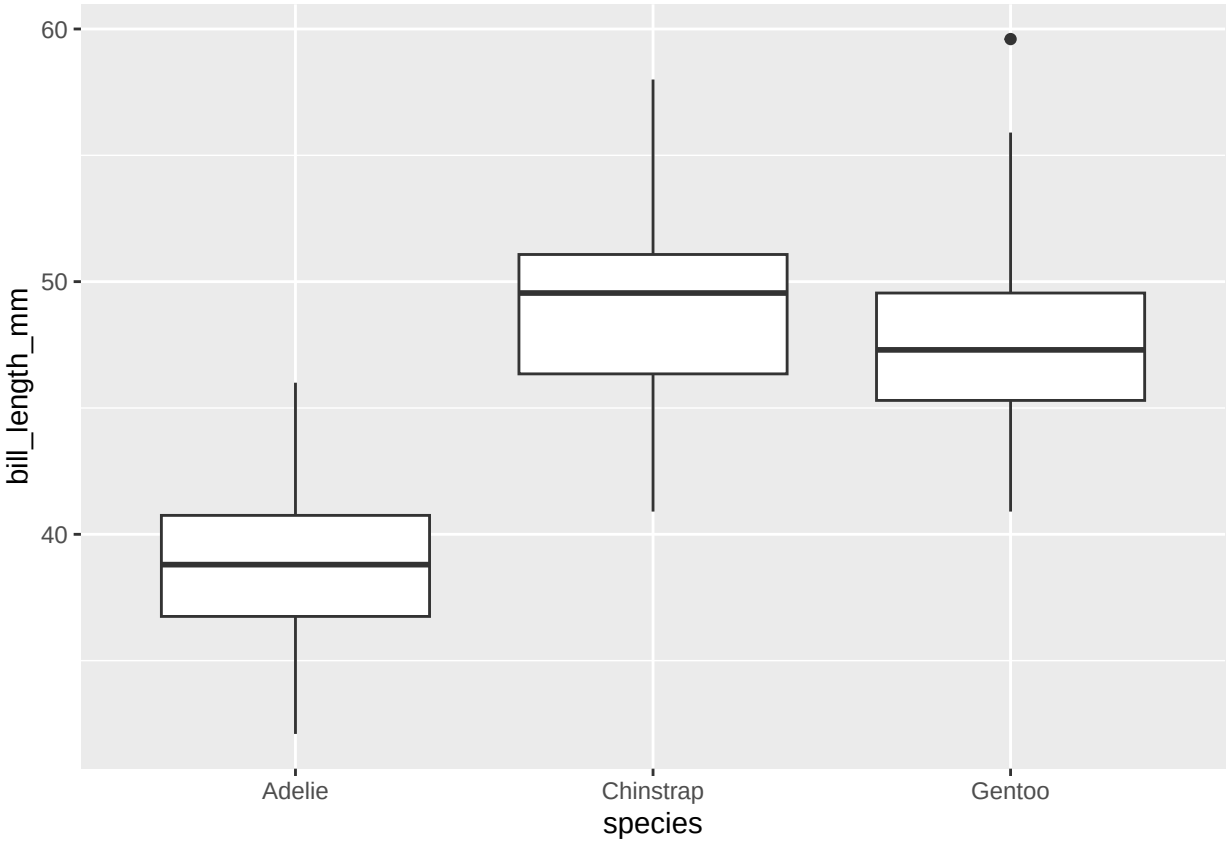
```
## Rows: 344
## Columns: 10
## $ X                    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15~
## $ species              <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, A~
## $ island               <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen~
## $ year                 <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2~
## $ bill_length_mm       <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.~
## $ bill_depth_mm        <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.~
## $ flipper_length_mm    <chr> "181", "186", "195", "NA", "193", "190", "181", "~
## $ body_mass_g          <chr> "3750", "3800", "3250", "NA", "3450", "3650", "36~
## $ above_average_weight <chr> "0", "0", "0", "NA", "0", "0", "0", "1", "0", "1"~
## $ sex                  <chr> "male", "female", "female", "NA", "female", "male~
```

Data Visualizations with transformed penguins dataframe

Boxplot

```
# Boxplot of bill lengths for each species type (boxplot plots the 5 number summary)
ggplot(
  data=penguins,
  aes(x=species,
      y=bill_length_mm)
) +
geom_boxplot()
```

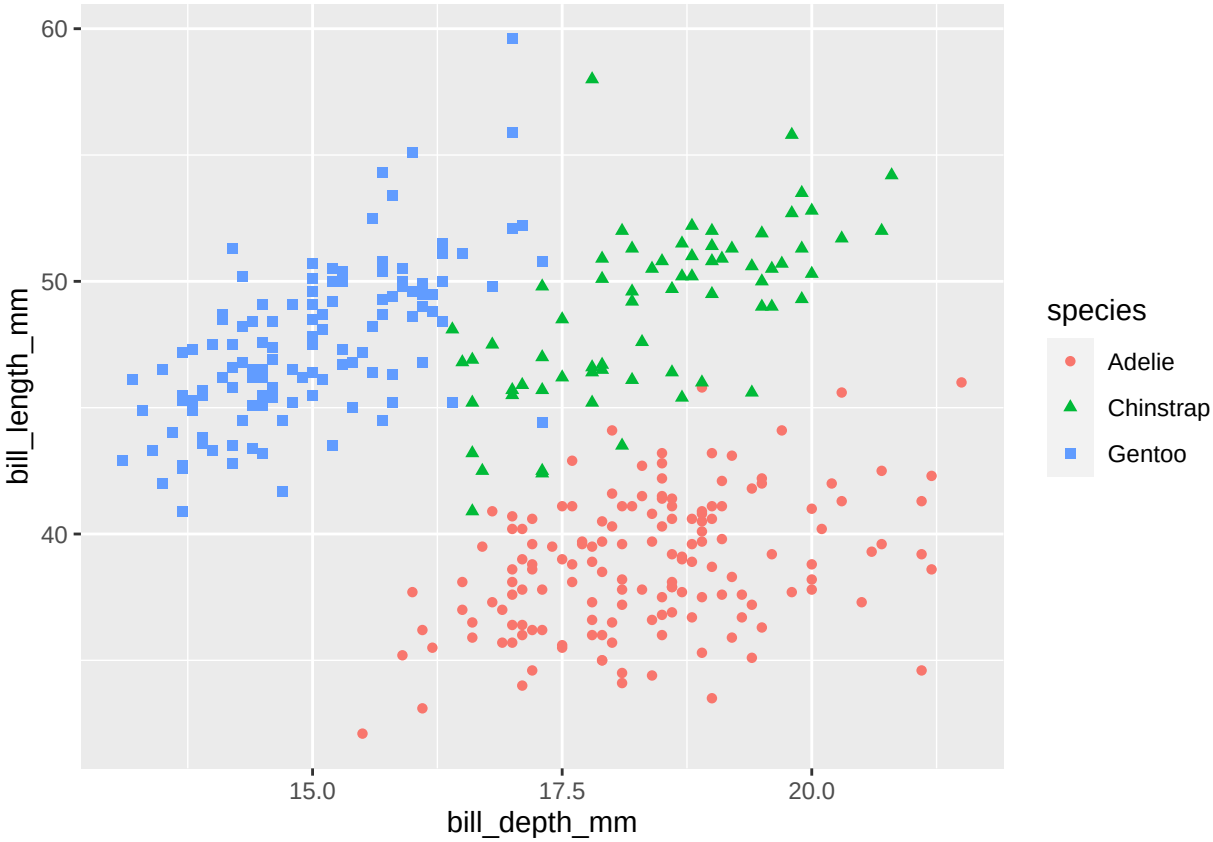
```
## Warning: Removed 2 rows containing non-finite values (`stat_boxplot()`).
```



Scatterplot

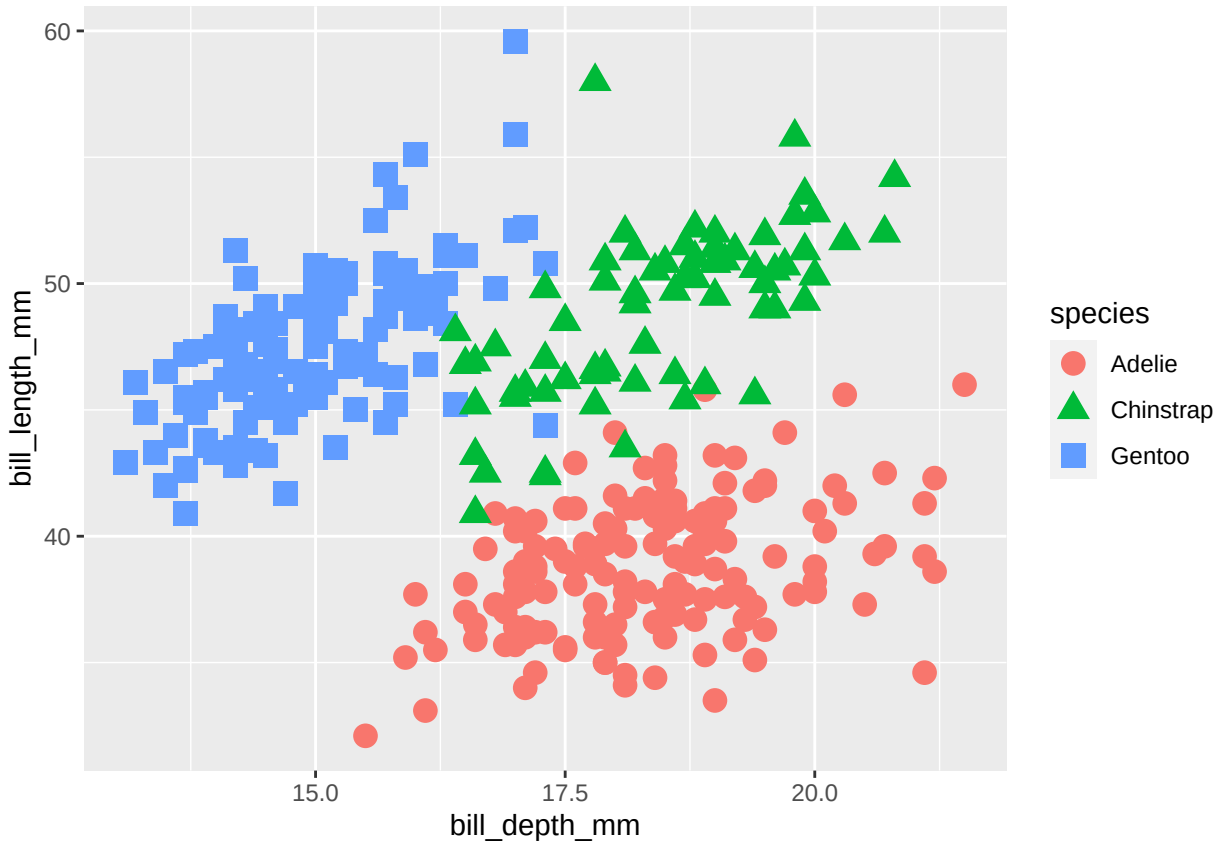
```
# Scatterplot of bill depth and bill length, marked by each species
ggplot(
  data=penguins,
  aes(x=bill_depth_mm,
      y=bill_length_mm,
      shape=species,
      color=species)
) +
geom_point()
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```




```
# Making mark of each species larger in previous plot
ggplot(
  data=penguins,
  aes(x=bill_depth_mm,
      y=bill_length_mm,
      shape=species,
      color=species)
) +
geom_point(size=4)
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```


60
50
40
bill_length_mm
15.0
17.5
20.0
bill_depth_mm
species
Adelie
Chinstrap
Gentoo

The tilde (~) and tilde dot (~. or .~) operators

The variable on the **left-hand side of tilde operator is the dependent variable** and the variable(s) on the **right-hand side of tilde operator is/are called the independent variable(s)**.

For example: Regression_Model <- lm(y~ x1 + x2 + x3)

If want to include all of a dataframe's columns, then can combine the tild operator with a dot (.).

For example: Regression_Model_New < - lm(y~ . , data = Regression_Data)

Stacking scatterplots horizontally

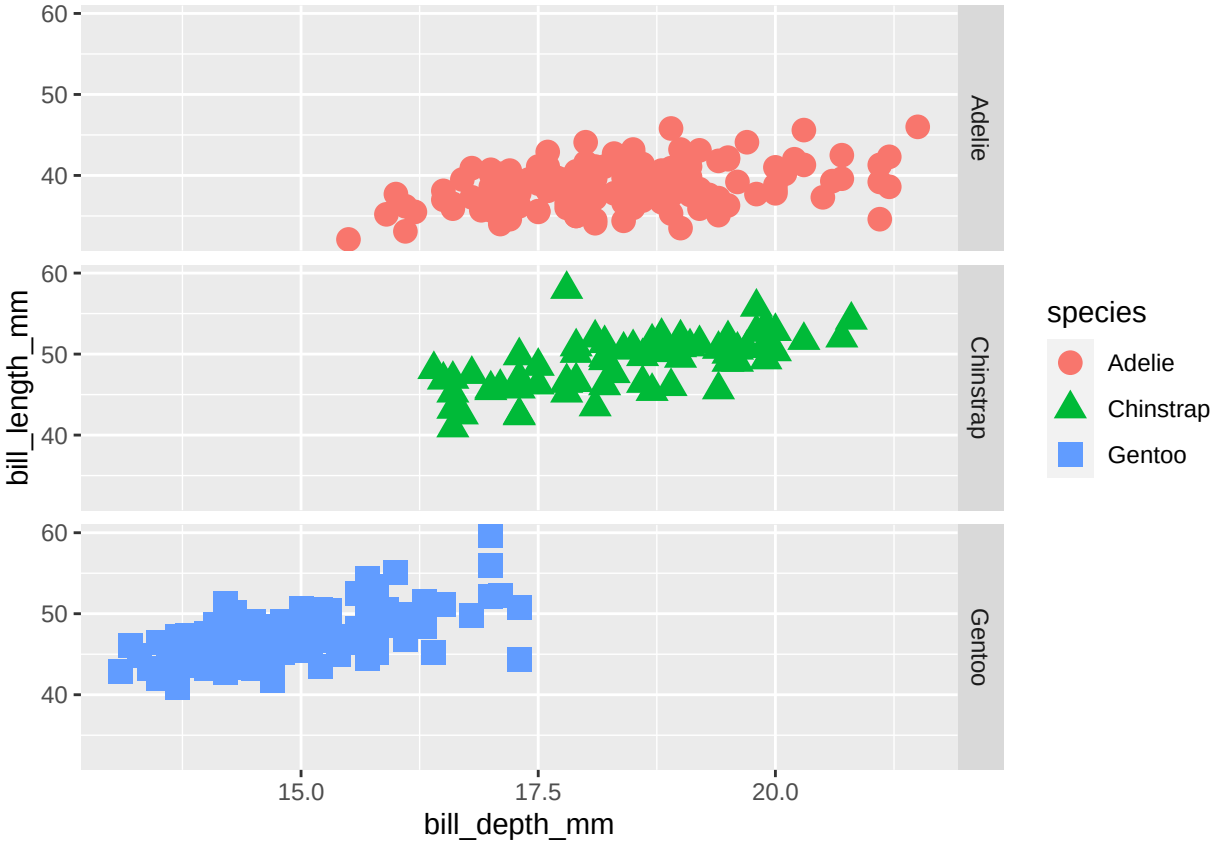
```
# Stack the scatterplots horizontally
ggplot(
  data=penguins,
  aes(x=bill_depth_mm,
      y=bill_length_mm,
      shape=species,
      color=species)
) +
geom_point(size=4) + facet_grid(.~species)
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

Stacking scatterplots vertically

```
ggplot(
  data=penguins,
  aes(x=bill_depth_mm,
      y=bill_length_mm,
      shape=species,
      color=species)
) +
geom_point(size=4) + facet_grid(species~.)
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



Dataframe methods & Chaining methods [LAPD dataset]

loading in dataset

```
lapd_file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/Discussion/R_Basics/Police_Payroll.csv
lapd = read.csv(lapd_file_path, na.strings='')

glimpse(lapd)
```

```
## Rows: 68,564
## Columns: 35
## $ Row.ID                          <chr> "3-1000027830ctFu", "3-1000155488ctFu", "~
## $ Year                            <int> 2013, 2013, 2013, 2013, 2013, 2013, 2013,~
## $ Department.Title                <chr> "Police (LAPD)", "Police (LAPD)", "Police~
## $ Payroll.Department              <int> 4301, 4302, 4301, 4301, 4302, 4302, 4301,~
## $ Record.Number                   <dbl> 1000027830, 1000155488, 1000194958, 10002~
## $ Job.Class.Title                 <chr> "Police Detective II", "Clerk Typist", "P~
## $ Employment.Type                 <chr> "Full Time", "Full Time", "Full Time", "F~
## $ Hourly.or.Event.Rate            <dbl> 53.16, 23.77, 60.80, 60.98, 45.06, 34.42,~
## $ Projected.Annual.Salary         <dbl> 110998.08, 49623.67, 126950.40, 127326.24~
## $ Q1.Payments                     <dbl> 24931.20, 11343.96, 24184.00, 29391.20, 2~
## $ Q2.Payments                     <dbl> 29181.61, 13212.37, 28327.20, 36591.20, 2~
## $ Q3.Payments                     <dbl> 26545.80, 11508.36, 28744.20, 32904.81, 2~
## $ Q4.Payments                     <dbl> 29605.30, 13442.53, 33224.88, 37234.03, 2~
## $ Payments.Over.Base.Pay          <dbl> 4499.12, 1844.82, 13192.43, 18034.53, 137~
## $ X..Over.Base.Pay                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Total.Payments                  <dbl> 110263.91, 49507.22, 114480.28, 136121.24~
## $ Base.Pay                        <dbl> 105764.79, 47662.40, 101287.85, 118086.71~
## $ Permanent.Bonus.Pay             <dbl> 3174.12, 0.00, 7363.95, 7086.67, 0.00, 0.~
## $ Longevity.Bonus.Pay             <dbl> 0.00, 1310.82, 0.00, 0.00, 1251.19, 1726.~
## $ Temporary.Bonus.Pay             <dbl> 1325.00, 0.00, 1205.00, 1325.00, 125.00, ~
## $ Lump.Sum.Pay                    <dbl> 0.00, 0.00, 2133.18, 0.00, 2068.80, 0.00,~
## $ Overtime.Pay                    <dbl> 0.00, 0.00, 4424.32, 9839.33, 0.00, 0.00,~
## $ Other.Pay...Adjustments         <dbl> 0.00, 534.00, -1934.02, -216.47, -2068.80~
## $ Other.Pay..Payroll.Explorer.    <dbl> 4499.12, 1844.82, 8768.11, 8195.20, 1376.~
## $ MOU                             <int> 24, 3, 24, 24, 12, 3, 24, 24, 24, 24, 24,~
## $ MOU.Title                       <chr> "POLICE OFFICERS UNIT", "CLERICAL UNIT", ~
## $ FMS.Department                  <int> 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 7~
## $ Job.Class                       <int> 2223, 1358, 2227, 2232, 1839, 2207, 2214,~
## $ Pay.Grade                       <chr> "2", "0", "1", "1", "0", "2", "3", "1", "~
## $ Average.Health.Cost             <dbl> 11651.40, 10710.24, 11651.40, 11651.40, 1~
## $ Average.Dental.Cost             <dbl> 898.08, 405.24, 898.08, 898.08, 405.24, 4~
## $ Average.Basic.Life              <dbl> 191.04, 11.40, 191.04, 191.04, 11.40, 11.~
## $ Average.Benefit.Cost            <dbl> 12740.52, 11126.88, 12740.52, 12740.52, 1~
## $ Benefits.Plan                   <chr> "Police", "City", "Police", "Police", "Ci~
## $ Job.Class.Link                  <chr> "http://per.lacity.org/perspecs/2223.pdf"~
```

Cleaning dataframe column names

```
lapd = clean_names(lapd)
glimpse(lapd)
```

```
## Rows: 68,564
## Columns: 35
## $ row_id                      <chr> "3-1000027830ctFu", "3-1000155488ctFu", "3-~
## $ year                        <int> 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2~
## $ department_title            <chr> "Police (LAPD)", "Police (LAPD)", "Police (~
## $ payroll_department          <int> 4301, 4302, 4301, 4301, 4302, 4302, 4301, 4~
## $ record_number               <dbl> 1000027830, 1000155488, 1000194958, 1000232~
## $ job_class_title             <chr> "Police Detective II", "Clerk Typist", "Pol~
## $ employment_type             <chr> "Full Time", "Full Time", "Full Time", "Ful~
## $ hourly_or_event_rate        <dbl> 53.16, 23.77, 60.80, 60.98, 45.06, 34.42, 4~
## $ projected_annual_salary     <dbl> 110998.08, 49623.67, 126950.40, 127326.24, ~
## $ q1_payments                 <dbl> 24931.20, 11343.96, 24184.00, 29391.20, 208~
## $ q2_payments                 <dbl> 29181.61, 13212.37, 28327.20, 36591.20, 241~
## $ q3_payments                 <dbl> 26545.80, 11508.36, 28744.20, 32904.81, 215~
## $ q4_payments                 <dbl> 29605.30, 13442.53, 33224.88, 37234.03, 252~
## $ payments_over_base_pay      <dbl> 4499.12, 1844.82, 13192.43, 18034.53, 1376.~
## $ x_over_base_pay             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ total_payments              <dbl> 110263.91, 49507.22, 114480.28, 136121.24, ~
## $ base_pay                    <dbl> 105764.79, 47662.40, 101287.85, 118086.71, ~
## $ permanent_bonus_pay         <dbl> 3174.12, 0.00, 7363.95, 7086.67, 0.00, 0.00~
## $ longevity_bonus_pay         <dbl> 0.00, 1310.82, 0.00, 0.00, 1251.19, 1726.16~
## $ temporary_bonus_pay         <dbl> 1325.00, 0.00, 1205.00, 1325.00, 125.00, 68~
## $ lump_sum_pay                <dbl> 0.00, 0.00, 2133.18, 0.00, 2068.80, 0.00, 0~
## $ overtime_pay                <dbl> 0.00, 0.00, 4424.32, 9839.33, 0.00, 0.00, 4~
## $ other_pay_adjustments       <dbl> 0.00, 534.00, -1934.02, -216.47, -2068.80, ~
## $ other_pay_payroll_explorer <dbl> 4499.12, 1844.82, 8768.11, 8195.20, 1376.19~
## $ mou                         <int> 24, 3, 24, 24, 12, 3, 24, 24, 24, 24, 24, 1~
## $ mou_title                   <chr> "POLICE OFFICERS UNIT", "CLERICAL UNIT", "P~
## $ fms_department              <int> 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70,~
## $ job_class                   <int> 2223, 1358, 2227, 2232, 1839, 2207, 2214, 2~
## $ pay_grade                   <chr> "2", "0", "1", "1", "0", "2", "3", "1", "B"~
## $ average_health_cost         <dbl> 11651.40, 10710.24, 11651.40, 11651.40, 107~
## $ average_dental_cost         <dbl> 898.08, 405.24, 898.08, 898.08, 405.24, 405~
## $ average_basic_life          <dbl> 191.04, 11.40, 191.04, 191.04, 11.40, 11.40~
## $ average_benefit_cost        <dbl> 12740.52, 11126.88, 12740.52, 12740.52, 111~
## $ benefits_plan               <chr> "Police", "City", "Police", "Police", "City~
## $ job_class_link              <chr> "http://per.lacity.org/perspecs/2223.pdf", ~
```

head() method

```
# We do head() to view only the first few lines of the data
head(select(lapd, year, base_pay))
```

```
##   year  base_pay
## 1 2013 105764.79
## 2 2013  47662.40
## 3 2013 101287.85
## 4 2013 118086.71
## 5 2013  90321.86
## 6 2013  62770.40
```

Selecting specific columns of a dataframe

```
lapd %>%
 select(year, base_pay)
```

Removing specific columns of a dataframe

```
# selects all of the columns except the ones with negative sign inside select() method
head(select(lapd, -row_id, -department_title))
```

```
##   year payroll_department record_number                job_class_title
## 1 2013               4301    1000027830             Police Detective II
## 2 2013               4302    1000155488                    Clerk Typist
## 3 2013               4301    1000194958              Police Sergeant I
## 4 2013               4301    1000232317             Police Lieutenant I
## 5 2013               4302    1000329284           Principal Storekeeper
## 6 2013               4302    1001124320 Police Service Representative II
##   employment_type hourly_or_event_rate projected_annual_salary q1_payments
## 1       Full Time                53.16               110998.08    24931.20
## 2       Full Time                23.77                49623.67    11343.96
## 3       Full Time                60.80               126950.40    24184.00
## 4       Full Time                60.98               127326.24    29391.20
## 5       Full Time                45.06                94076.67    20813.00
## 6       Full Time                34.42                71871.57    16056.87
##   q2_payments q3_payments q4_payments payments_over_base_pay x_over_base_pay
## 1    29181.61    26545.80    29605.30                4499.12               0
## 2    13212.37    11508.36    13442.53                1844.82               0
## 3    28327.20    28744.20    33224.88               13192.43               0
## 4    36591.20    32904.81    37234.03               18034.53               0
## 5    24136.00    21517.76    25231.29                1376.19               0
## 6    17926.86    14149.91    17051.92                2415.16               0
##   total_payments  base_pay permanent_bonus_pay longevity_bonus_pay
## 1      110263.91 105764.79             3174.12               0.00
## 2       49507.22  47662.40                0.00            1310.82
## 3      114480.28 101287.85             7363.95               0.00
## 4      136121.24 118086.71             7086.67               0.00
## 5       91698.05  90321.86                0.00            1251.19
## 6       65185.56  62770.40                0.00            1726.16
##   temporary_bonus_pay lump_sum_pay overtime_pay other_pay_adjustments
## 1                1325         0.00         0.00                  0.00
```

```
## 2                   0         0.00         0.00               534.00
## 3                1205      2133.18      4424.32             -1934.02
## 4                1325         0.00      9839.33              -216.47
## 5                 125      2068.80         0.00             -2068.80
## 6                 689         0.00         0.00                0.00
##   other_pay_payroll_explorer mou           mou_title fms_department job_class
## 1                    4499.12  24 POLICE OFFICERS UNIT             70      2223
## 2                    1844.82   3        CLERICAL UNIT             70      1358
## 3                    8768.11  24 POLICE OFFICERS UNIT             70      2227
## 4                    8195.20  24 POLICE OFFICERS UNIT             70      2232
## 5                    1376.19  12     SUPV BLUE COLLAR             70      1839
## 6                    2415.16   3        CLERICAL UNIT             70      2207
##   pay_grade average_health_cost average_dental_cost average_basic_life
## 1         2            11651.40              898.08             191.04
## 2         0            10710.24              405.24              11.40
## 3         1            11651.40              898.08             191.04
## 4         1            11651.40              898.08             191.04
## 5         0            10710.24              405.24              11.40
## 6         2            10710.24              405.24              11.40
##   average_benefit_cost benefits_plan                          job_class_link
## 1             12740.52        Police http://per.lacity.org/perspecs/2223.pdf
## 2             11126.88          City http://per.lacity.org/perspecs/1358.pdf
## 3             12740.52        Police http://per.lacity.org/perspecs/2227.pdf
## 4             12740.52        Police http://per.lacity.org/perspecs/2232.pdf
## 5             11126.88          City http://per.lacity.org/perspecs/1839.pdf
## 6             11126.88          City http://per.lacity.org/perspecs/2207.pdf
```

starts_with() method

```
head(select(lapd, starts_with('q')))
```

```
##   q1_payments q2_payments q3_payments q4_payments
## 1    24931.20    29181.61    26545.80    29605.30
## 2    11343.96    13212.37    11508.36    13442.53
## 3    24184.00    28327.20    28744.20    33224.88
## 4    29391.20    36591.20    32904.81    37234.03
## 5    20813.00    24136.00    21517.76    25231.29
## 6    16056.87    17926.86    14149.91    17051.92
```

ends_with() method

```
head(select(lapd, ends_with('pay')))
```

```
##   payments_over_base_pay x_over_base_pay  base_pay permanent_bonus_pay
## 1                4499.12               0 105764.79             3174.12
## 2                1844.82               0  47662.40                0.00
## 3               13192.43               0 101287.85             7363.95
## 4               18034.53               0 118086.71             7086.67
## 5                1376.19               0  90321.86                0.00
## 6                2415.16               0  62770.40                0.00
##   longevity_bonus_pay temporary_bonus_pay lump_sum_pay overtime_pay
## 1                0.00                1325         0.00         0.00
## 2             1310.82                   0         0.00         0.00
## 3                0.00                1205      2133.18      4424.32
## 4                0.00                1325         0.00      9839.33
## 5             1251.19                 125      2068.80         0.00
## 6             1726.16                 689         0.00         0.00
```

Slicing - retrieving specific rows

```
slice(lapd, 3:7)
```

```
##             row_id year department_title payroll_department record_number
## 1 3-1000194958ctFu 2013    Police (LAPD)               4301    1000194958
## 2 3-1000232317ctFu 2013    Police (LAPD)               4301    1000232317
## 3 3-1000329284ctFu 2013    Police (LAPD)               4302    1000329284
## 4 3-1001124320ctFu 2013    Police (LAPD)               4302    1001124320
## 5 3-1001221822ctFu 2013    Police (LAPD)               4301    1001221822
##                    job_class_title employment_type hourly_or_event_rate
## 1                Police Sergeant I       Full Time                60.80
## 2              Police Lieutenant I       Full Time                60.98
## 3            Principal Storekeeper       Full Time                45.06
## 4 Police Service Representative II       Full Time                34.42
## 5               Police Officer III       Full Time                47.76
##   projected_annual_salary q1_payments q2_payments q3_payments q4_payments
## 1               126950.40    24184.00    28327.20    28744.20    33224.88
## 2               127326.24    29391.20    36591.20    32904.81    37234.03
## 3                94076.67    20813.00    24136.00    21517.76    25231.29
## 4                71871.57    16056.87    17926.86    14149.91    17051.92
## 5                99722.88    22162.22    25664.40    23404.40    24586.05
##   payments_over_base_pay x_over_base_pay total_payments  base_pay
## 1               13192.43              0      114480.28 101287.85
## 2               18034.53              0      136121.24 118086.71
## 3                1376.19              0       91698.05  90321.86
## 4                2415.16              0       65185.56  62770.40
## 5                2099.31              0       95817.07  93717.76
##   permanent_bonus_pay longevity_bonus_pay temporary_bonus_pay lump_sum_pay
## 1             7363.95               0.00                1205      2133.18
## 2             7086.67               0.00                1325         0.00
## 3                0.00            1251.19                 125      2068.80
## 4                0.00            1726.16                 689         0.00
## 5              866.27               0.00                1145         0.00
##   overtime_pay other_pay_adjustments other_pay_payroll_explorer mou
## 1      4424.32              -1934.02                    8768.11  24
## 2      9839.33               -216.47                    8195.20  24
## 3         0.00              -2068.80                    1376.19  12
## 4         0.00                  0.00                    2415.16   3
## 5        42.14                 45.90                    2057.17  24
##              mou_title fms_department job_class pay_grade average_health_cost
## 1 POLICE OFFICERS UNIT             70      2227         1            11651.40
## 2 POLICE OFFICERS UNIT             70      2232         1            11651.40
## 3     SUPV BLUE COLLAR             70      1839         0            10710.24
## 4        CLERICAL UNIT             70      2207         2            10710.24
## 5 POLICE OFFICERS UNIT             70      2214         3            11651.40
##   average_dental_cost average_basic_life average_benefit_cost benefits_plan
## 1              898.08             191.04              12740.52        Police
## 2              898.08             191.04              12740.52        Police
## 3              405.24              11.40              11126.88          City
## 4              405.24              11.40              11126.88          City
## 5              898.08             191.04              12740.52        Police
##                            job_class_link
## 1 http://per.lacity.org/perspecs/2227.pdf
```

```
## 2 http://per.lacity.org/perspecs/2232.pdf
## 3 http://per.lacity.org/perspecs/1839.pdf
## 4 http://per.lacity.org/perspecs/2207.pdf
## 5 http://per.lacity.org/perspecs/2214.pdf
```

Filerting - filer columns based on boolean predicate

```
filter(lapd, year == 2018)
```

Method chaining

```
# filter lapd on year = 2018 and base_pay > 62474
lapd %>%
  filter(year == 2018 & base_pay > 62474)

# filter lapd on year = 2018 and base_pay > 62474 and display the row counts
lapd %>%
  filter(year == 2018 & base_pay > 62474) %>%
  nrow()

# filter lapd on year >= 2013 and year <= 2015 and display the row counts
lapd %>%
  filter(year >= 2013 & year <= 2015) %>%
  nrow()

# filter lapd on employment_type = full time and year = 2018 and display the row counts
lapd %>%
  filter(employment_type == "Full Time" & year == 2018) %>%
  nrow()
```

Method chaining - 2 chains

```
# apply a filter to lapd and then select specific columns of the filtered dataframe
lapd %>%
  filter(year == 2018) %>%
  select(job_class_title,
         employment_type,
         base_pay)
```

Dataframe transformation of data with mutate() method

```
# divide every base_pay_k value by 1000
lapd %>%
  mutate(base_pay_k = base_pay/1000)
```

View summary of mutated lapd dataframe

```
summarize(lapd, mean_base_pay = mean(base_pay))
```

```
##   mean_base_pay
## 1      84789.67
```

Creating a new column with boolean logic

```
# Create a new variable called base_pay_level which has Less Than 0, No Income, Less than Median and Gr
lapd %>%
  mutate(base_pay_level = case_when(
    base_pay < 0 ~ "Less than 0",
    base_pay == 0 ~ "No Income",
    base_pay < 62474 & base_pay > 0 ~ "Less than Median, Greater than 0",
    base_pay > 62474 ~ "Greater than Median"))
```

Aggregate dataframe data using group_by() method

```
lapd %>%
  group_by(employment_type) %>%
  summarize(med_base_pay = median(base_pay),
            count = n())
```

```
## # A tibble: 3 x 3
##   employment_type med_base_pay count
##   <chr>                  <dbl> <int>
## 1 Full Time             90943. 67872
## 2 Part Time             17530.   521
## 3 Per Event              4950    171
```