APSTA Week 02 Exercise Solutions

1. The data set `pi2000` (UsingR) contains the first 2,000 digits of π. What is the percentage of digits that are 3 or less? What is the percentage of the digits are 5 or more?
Solution:

```
library(UsingR)
three_leq <- length(pi2000[pi2000 <= 3])
five_geq <- length(pi2000[pi2000 >= 5])
sprintf(
  "Around %s%% of the digits are three or less, while around %s%% are five or more.",
  round(
    (three_leq/2000)*100, 3
  ),
  round(
    (five_geq/2000)*100, 3
  )
)
```

```
## [1] "Around 39.5% of the digits are three or less, while around 50.75% are five or more."
```

2. Load the `starwars` dataset from the `dplyr` package. The dataset has 87 characters from the StarWars universe with 13 features. **Note** there are missing data in the database represented by `NA` values.

```
starwars <- dplyr::starwars
```

a) What is the homeworld of Mace Windu?
Solution:

```
sprintf("Mase Windu is from %s", starwars[starwars$name == "Mace Windu",]$homeworld)
```

```
## [1] "Mase Windu is from Haruun Kal"
```

b) How many droids are in the dataset?
Solution:

```
sprintf(
  "The dataset includes %s different droids.",
  nrow(starwars[starwars$species == "Droid",])
)
```

```
## [1] "The dataset includes 10 different droids."
```

c) Who are the shortest and tallest humans in dataset?
Solution:

```r
min_height <- min(starwars$height, na.rm=TRUE)
max_height <- max(starwars$height, na.rm=TRUE)
shortest <- starwars[starwars$height == min_height,][1,]$name
tallest <- starwars[starwars$height == max_height & !is.na(starwars$height),]$name
sprintf(
  "Smallest person in Star Wars: %s. Tallest person in Star Wars: %s.",
  shortest, tallest
)
```

```
## [1] "Smallest person in Star Wars: Yoda. Tallest person in Star Wars: Yarael Poof."
```

d) What is the mean and standard deviation of the height all humans in the starwars database?

```r
mean_height <- mean(starwars[starwars$species == "Human",]$height, na.rm=TRUE)
SD_height <- sd(starwars[starwars$species == "Human",]$height, na.rm=TRUE)
sprintf("The mean height of Humans in Star Wars is %scm with a standard deviation of %s.",
        round(mean_height, 1), round(SD_height, 1))
```

```
## [1] "The mean height of Humans in Star Wars is 176.6cm with a standard deviation of 12.5."
```

3. The following table shows the results of a survey of 10 pirates. In addition to some basic demographic information, the survey asked each pirate "What is your favorite superhero?"" and "How many tattoos do you have?""

Name	Sex	Age	Superhero	Tattoos
Astrid	F	30	Batman	11
Lea	F	25	Superman	15
Sarina	F	25	Batman	12
Remon	M	29	Spiderman	5
Letizia	F	22	Batman	65
Babice	F	22	Antman	3
Jonas	M	35	Batman	9
Wendy	F	19	Superman	13
Niveditha	F	32	Maggott	900
Gioia	F	21	Superman	0

a) Combine the data into a single dataframe.
Complete all the following exercises from the dataframe!
Solution:

```r
pirates <- data.frame(
  name = c("Astrid", "Lea", "Sarina", "Remon", "Letizia", "Babice",
           "Jonas", "Wendy", "Niveditha", "Gioia"),
  sex = c("F", "F", "F", "M", "F", "F", "M", "F", "F", "F"),
  age = c(30, 25, 25, 29, 22, 22, 35, 19, 32, 21),
  superhero = c("Batman", "Superman", "Batman", "Spiderman", "Batman",
                "Antman", "Batman", "Superman", "Maggott", "Superman"),
  tattoos = c(11, 15, 12, 5, 65, 3, 9, 13, 900, 0)
)
head(pirates)
```

```
##      name sex age superhero tattoos
## 1  Astrid   F  30    Batman      11
## 2     Lea   F  25  Superman      15
## 3  Sarina   F  25    Batman      12
## 4   Remon   M  29 Spiderman       5
## 5 Letizia   F  22    Batman      65
## 6  Babice   F  22    Antman       3
```

b) What was the mean age of female and male pirates separately?
Solution:

```
male_mean <- mean(pirates[pirates$sex == "M",]$age)
female_mean <- mean(pirates[pirates$sex == "F",]$age)
sprintf("Male mean: %s. Female mean: %s.", male_mean, female_mean)
```

```
## [1] "Male mean: 32. Female mean: 24.5."
```

c) Add a new column to the dataframe called `tattoos.per.year` which shows how many tattoos each pirate has for each year in their life. Which pirate had the most number of tattoos per year?
Solution:

```
pirates$tattoos.per.year <- pirates$tattoos / pirates$age
tattoo_freak = pirates[pirates$tattoos.per.year == max(pirates$tattoos.per.year),][1,]
sprintf("%s had %s tattoos by the age of %s. That's a whopping %s tattoos per year!",
        tattoo_freak$name,
        tattoo_freak$tattoos,
        tattoo_freak$age,
        tattoo_freak$tattoos.per.year)
```

```
## [1] "Niveditha had 900 tattoos by the age of 32. That's a whopping 28.125 tattoos per year!"
```

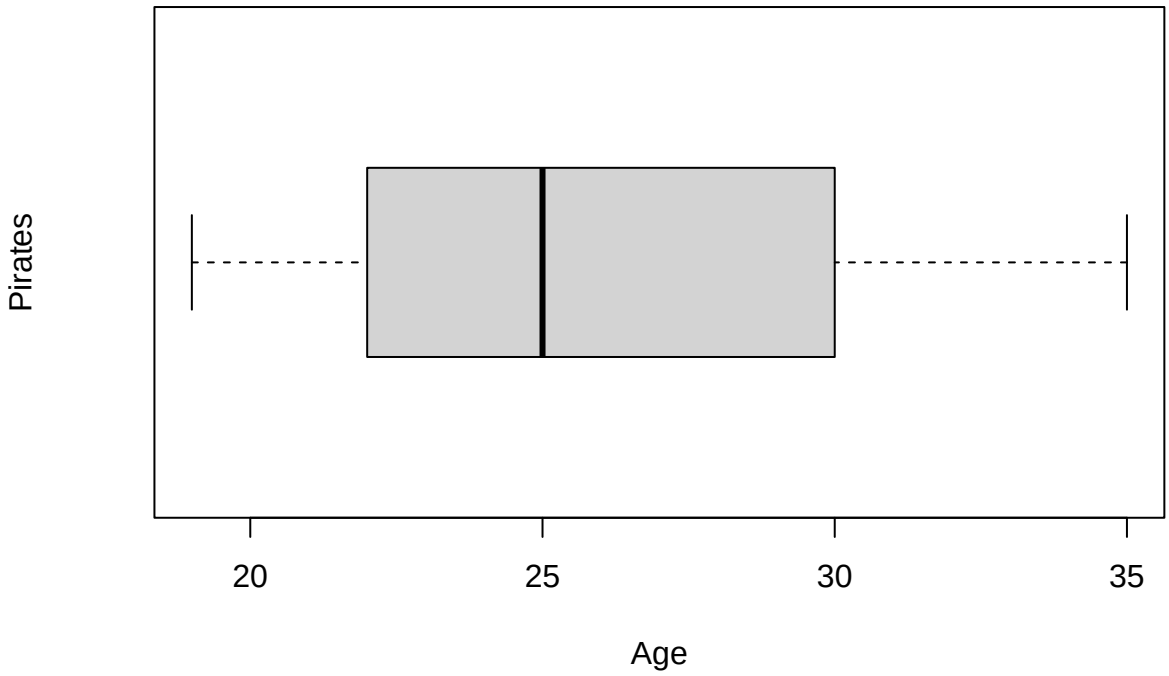
d) What was the median number of tattoos of pirates over the age of 20 whose favorite superhero is Spiderman?
Solution:

```
filtered <- pirates[
  pirates$age > 20 &
  pirates$superhero == "Spiderman",
]
sprintf(
  "Median tattoos: %s (though there is only one person who fits the criteria).",
  median(filtered$tattoos)
)
```

```
## [1] "Median tattoos: 5 (though there is only one person who fits the criteria)."
```

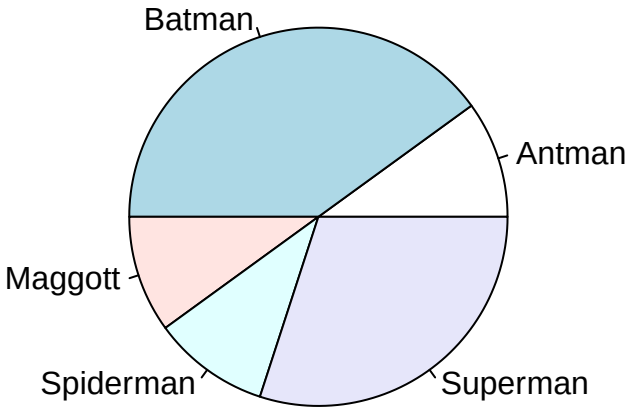
e) Make a boxplot of the age distribution of the pirates
Solution:

```
boxplot(
  pirates$age,
  xlab="Age", ylab="Pirates",
  horizontal=TRUE
)
```



f) Make a piechart showing the number of pirates which has each superhero as their favorite.
Solution:

```
pie(table(pirates$superhero))
```


Batman
Antman
Maggott
Spiderman
Superman