Seminar task and questions: Week 2 Solutions

Solutions for task

This is how to create an object with the "**Suburbs and Small Towns**" category by filtering it from the data frame using this code:

```
Subs_towns <- London.Ambulance[London.Ambulance$WardType == 'Suburbs and Small Towns',]
```

Computing the summary statistics for this category is as follows:

Calculate the **mode** using the created `get_mode()`.

```
# create a function to calculate the mode
get_mode <- function(x) {
    # get unique values of the input vector
    uniqv <- unique(x)
    # select the values with the highest number of occurrences
    uniqv[which.max(tabulate(match(x, uniqv)))]
}
# calculate the mode of the 2011 population variable
get_mode(Subs_towns$Assault_09_11)
```

```
## [1] 89
```

Obtain it's summary statistics for mean, median, 25th & 75th quartiles, and min and max values of the fly by using the `summary()` function:

```
summary(Subs_towns$Assault_09_11)
```

```
## Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 19.0    67.0    98.0   110.0   142.8   579.0
```

To calculate the standard deviation:

```
sd(Subs_towns$Assault_09_11)
```

```
## [1] 64.5271
```

Full interpretation: The overall mean number of reported incidents across 210 wards defined as **Suburbs and Small Towns** was 110.1 (with SD of ± 64.5271) with the following quartiles: Median = 98.0; IQRs 67 to 142.8. The lower observed number of incidents of assaults in a ward was 19 and the highest was 579. Note that from the Mode, the most frequent reported number of assaults is 84.

Solutions for **the questions**

This is how to create an object with the "Prospering Metropolitan" category by filtering it from the data frame using this code:

```
ProsMetro <- London.Ambulance[London.Ambulance$WardType == 'Prospering Metropolitan',]
```

We essentially repeating what we did for the task

Calculate the **mode** using the `get_mode()`.

```
# create a function to calculate the mode
get_mode <- function(x) {
    # get unique values of the input vector
    uniqv <- unique(x)
    # select the values with the highest number of occurrences
    uniqv[which.max(tabulate(match(x, uniqv)))]
}
# calculate the mode of the 2011 population variable
get_mode(ProsMetro$Assault_09_11)
```

```
## [1] 0
```

Obtain it's summary statistics for mean, median, 25th & 75th quartiles, and min and max values of the fly by using the `summary()` function:

```
summary(ProsMetro$Assault_09_11)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0    75.0   119.0   153.3   184.0  1582.0
```

To calculate the standard deviation:

```
sd(ProsMetro$Assault_09_11)
```

```
## [1] 179.8195
```

Differences

Comparartive analysis (always best to present it as a table)

	Suburbs & Small towns	**Prospering Metropolitan**
Summary estimate		
Number of wards (n)	210	169
Mean (±SD)	110.0 (±64.5)	153.3 (±179.8)
Median (IQR)	98 (67 to 143)	119 (75 to 184)
Minimum and Maximum	19 and 579	0 and 1582
Mode	84	0

Full interpretation: Based on the mean (or median) estimates, the wards classed as **Suburbs & Small towns** tend to report less assaults than **Prospering Metropolitan** because the mean for the former is lower than the latter (i.e., `110.0` < `153.3`). In terms of dispersion, the points are more clustered to the mean for **Suburbs & Small towns** since the standard deviation for this group is lower than **Prospering Metropolitan**. The large standard variation, including wide range values (i.e., minimum and maximum values) for **Prospering Metropolitan** (i.e., `1582-0 = 1582.0`) indicates greater spread in this group.

Code for creating the dual box plot

```
# To generate the dual boxplot use rbind() to combine the dfs

data <- rbind(Subs_towns, ProsMetro)

# dual boxplot :::: use option outline=FALSE to exclude outliers

boxplot(data$Assault_09_11 ~ data$WardType, outline = FALSE, xlab = "",
ylab="Levels of assault", main="Boxplot [Note: Outliers were excluded]")
```