Module: Advanced graphics

R program course
Jan-Feb-Mar 2018
CCHE -57357

EXERCISE

Advanced graphics (box-scatter)

Use the excel file "RCC" containing the data about Renal cell carcinoma patients

Draw a **boxplot** for ANXA, PDCD4 and DFFA through following the steps below

1. View the data to:
    a. identify the column numbers of the 3 variables above (from column no. …. to column no. ….)
    b. identify the column numbers containing gender, side and grade (from column no. …. to column no. ….)

2. Select from RCC the columns identified in step one genes' in a new file named 'genes'

3. Melt 'genes' but put "gender", "Side" and "Grade" as id variables.

    Name the new file 'genesmelted'. See how it looks like
    Make sure the spelling of the variables is correct

4. Now lets make a nice boxplot step by step

    a. Draw a boxplot for genesmelted where x=variable and y=value
        Remember, variable and value represent the 3 genes without the id variables
    Use the minimum required code to draw it i.e. only the necessary part (aes and geom_boxplot)
    It looks bad because the data has many outliers, however they are real data and we can't remove them, so let's do the next step

    b. Log transform the values for a better visualization
        Much better!
        Now, let's add another variable

c. Fill the colour of the boxplot according to Grade (fill=Grade in aes)
That looks nice but there are no spaces between the boxplots of each group. Let's fix that

d. Make a space between the boxes through
   i. Setting the position to position=position_dodge(0.6) in geom_boxplot change the number (0.6) to suite your graph
   ii. Making the box thinner through width in boxplot (eg.width=0.5) in geom_boxplot change the width to suite your graph

e. Make the outliers invisible (outlier.color = NA in geom_boxplot) and the background white (add theme_bw)

f. Make the y-axis title "Fold change (log transformed)" and remove the x-axis title (use x-lab and y-lab)

g. Increase the size of the x-axis labels (size=10) and make them bold through adding a theme(axis.text.x)

h. That looks good! Now, export your image

Draw a **boxplot** overlaid with dotplot for ANXA, PDCD4 and DFFA through following the steps below

1. Using the same code above, but
   a. Add the geom_dotplot
   b. Specify a fill for the boxplot as grey (fill="gray" in geom_boxplot
   Now the dots are filled according to grade instead of the boxes

2. Add a notch to the boxplot (notch =TRUE in geom_boxplot)

3. Fill the dot colour by gender instead of Grade (fill=gender instead of fill=Grade in aes)

4. Beautiful graph! Export it.

Draw a **scatter plot** showing the relation between E2F3 and MALAT1 depending on Side, gender and Grade through following the steps below

Now, remember, in scatter plots. You don't need to melt the data, so use the Original file 'RCC'

1. Make a very basic plot i.e only the necessary parts (aes and geom_point) where x=E2F3 and y= MALAT1

2. Make the dots coloured "gold" and half transparent (alpha=0.5) [in geom_point] and the background dark (add theme_dark)

3. Fill the points with different colours according to a new variable (Side) through

    a. Adding to aes, fill=Side
    b. Choosing a dot shape that is empty from inside, thus can be filled (eg: shape=21)

    Now you are displaying 3 variables in one plot; 2 quantitative variables (E2F3 and MALAT1) and one qualitative/grouping variable (Side)

4. Increase the dot size to 4 (size=4 in geom_point), remove the colour gold (color="gold" in geom_point) surrounding the dots

    That looks much better!

5. Add another variable 'Grade', make the size of the points vary according to the Grade through
    a. adding in aes, size=Grade
    b. removing the size part from the geom_point

Now you have 4 variables in the plot (2 quantitative and 2 grouping [Side and Grade])

6. Cool graph! Export