

QTM 151

Week 2 – ggplot2

Umberto Mignozzetti

Feb 05

Before we get started

- Did you set up your GitHub account?
- If not, please do it by next Monday.
- If you want, you can also request GitHub to get education discount: I have my GitHub for free since 2015!
- By Tuesday I'll post on Canvas how to connect R with Github.
- The quiz will be posted today, at 4:00PM
- Our GitHub page is:
<https://github.com/umbertomig/qtm151>

Today's Agenda

- Today we will learn `qplot` graphs.
- Graphs for Continuous Variables
- Graphs for Discrete Variables
- Graphs for Two Continuous Variables
- Graphs for Continuous x Discrete Variables

qplot

qplot

- `qplot`, as the name says, stands for quick plots.
- It is great to generate `ggplot` graphs in a hurry.
- And the graphs are indeed `ggplot` objects: you can still add layers to it!

qplot - Geoms Available

- Which graphs can we generate?
 - **"point"**: scatterplots.
 - **"line"**: line plot.
 - **"histogram"**
 - **"boxplot"**
 - **"density"**
 - **"bar"**: barplot.
 - **"smooth"**: Fits a smooth line.
 - **"dotplot"**: dotplot.

qplot - Options for Customization

- And there are plenty of quick options to customize the graphs.
 - `data`: Specify the data-frame.
 - `main`: Title.
 - `xlab`, `ylab`: x and y axis labels.
 - `color`: Controls the color of the lines/points.
 - `fill`: Controls the color of areas (e.g. for histograms).
 - `size`: Controls the size of points.
 - `shape`: The shape of points ("circle", "square", "triangle", etc...)
 - `alpha`: Controls the level of transparency of points/lines/fills.
 - `lwd`: Line width.
 - `lty`: Line type ("solid", "dashed", "dotted", etc...).
 - `facets`: Split up the data into multiple plots.

Questions?

Loading tidyverse

```
# Load tidyverse
```

```
library(tidyverse)
```

```
## — Attaching packages ————— tidyverse
```

```
## ✓ ggplot2 3.3.2      ✓ purrr 0.3.4
```

```
## ✓ tibble 3.0.4       ✓ dplyr 1.0.2
```

```
## ✓ tidyr 1.1.2        ✓ stringr 1.4.0
```

```
## ✓ readr 1.4.0        ✓ forcats 0.5.0
```

```
## — Conflicts ————— tidyverse_0.1.0
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

Loading datasets

```
# Loading tips dataset
```

```
tips ← read.csv('https://raw.githubusercontent.com/umbertomig/qtn  
head(tips, 2)
```

```
##      obs totbill  tip sex smoker day  time size  
## 1      1   16.99 1.01  F      No Sun Night    2  
## 2      2   10.34 1.66  M      No Sun Night    3
```

```
# Loading PErisk dataset
```

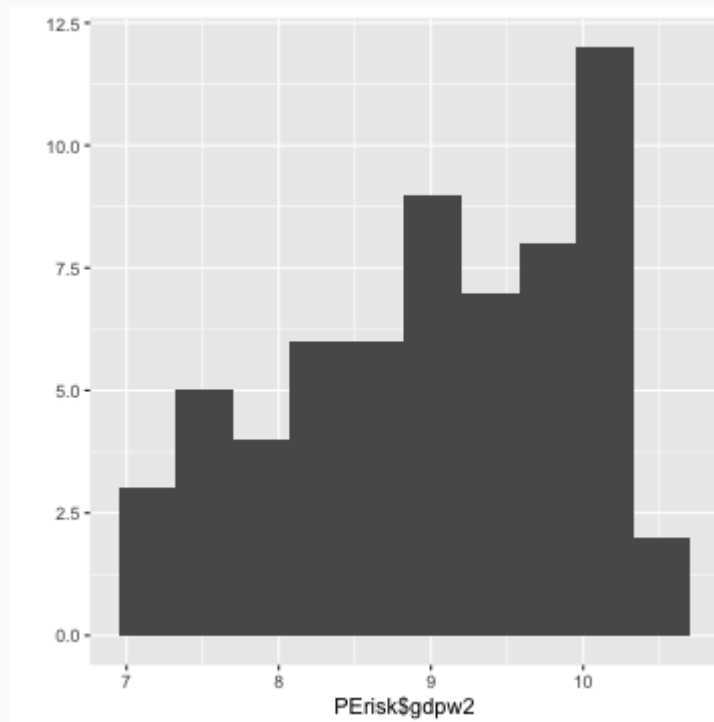
```
PErisk ← read.csv('https://raw.githubusercontent.com/umbertomig/c  
head(PErisk, 2)
```

```
##      country courts      barb2 prsexp2 prscorr2      gdpw2  
## 1 Argentina      0 -0.7207754      1      3  9.69017  
## 2 Australia      1 -6.9077550      5      4 10.30484
```

Plots for Continuous Variables

qplot - Histograms

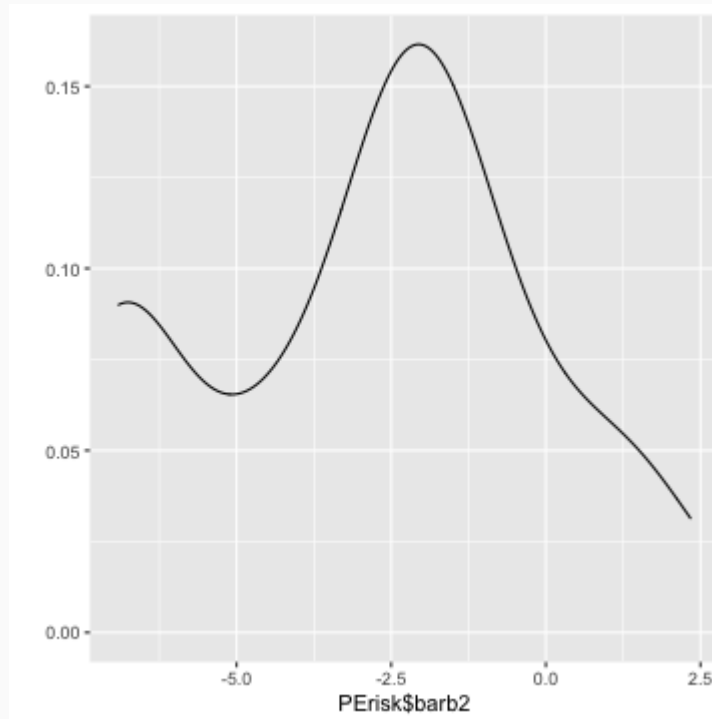
```
qplot(PErisk$gdpw2, geom = "histogram", bins = 10)
```



- **Your turn:** make a histogram of the `tip` variable in the `tips` dataset.

qplot - Density-plots

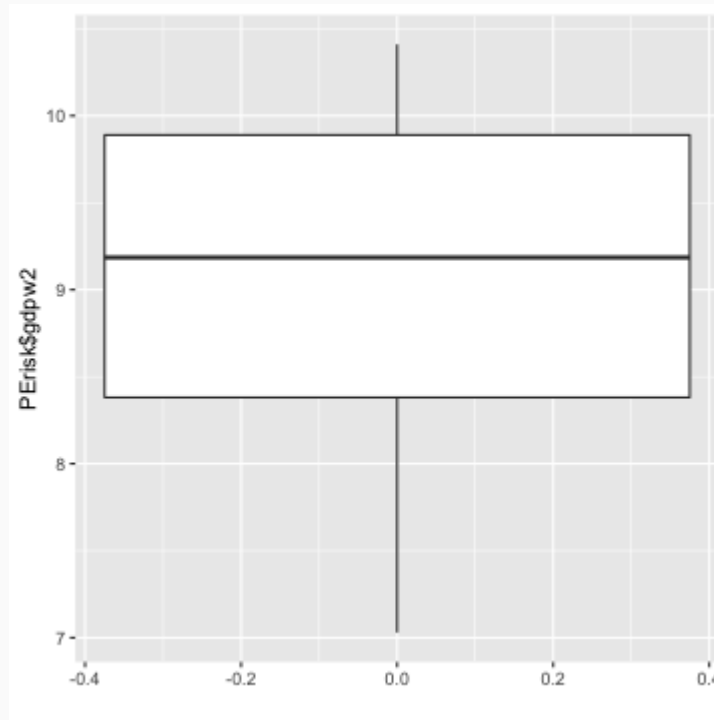
```
qplot(PERisk$barb2, geom = "density")
```



- **Your turn:** make a density plot of the `totbill` variable in the `tips` dataset.

qplot - Box-plots

```
# Box-plot of log of per capita gdp  
qplot(y = PErisk$gdpw2, geom = "boxplot")
```



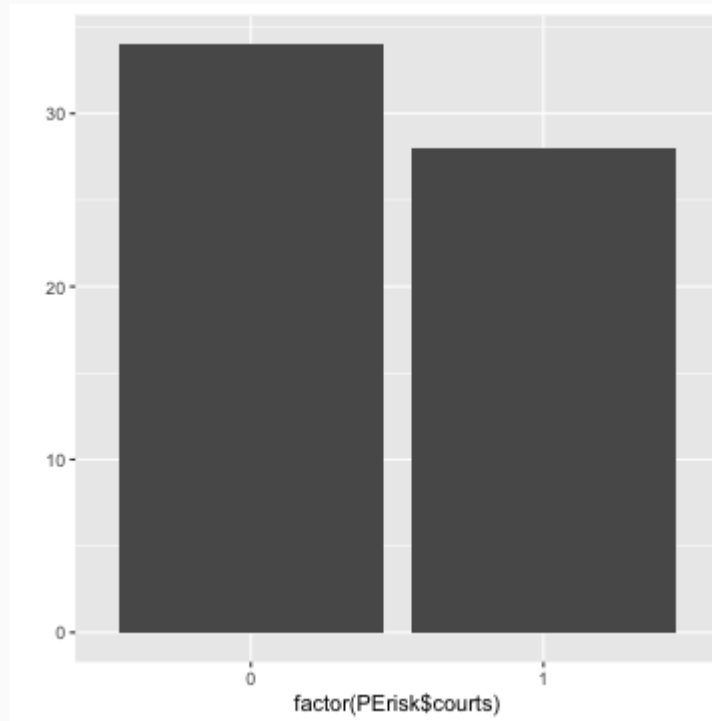
- **Your turn:** make a boxplot of the `totbill` variable in the `tips` dataset.

Plot for Discrete Variables

qplot - Bar-Plots

```
# Bar-plot of courts
```

```
qplot(factor(PERisk$courts), geom = "bar")
```

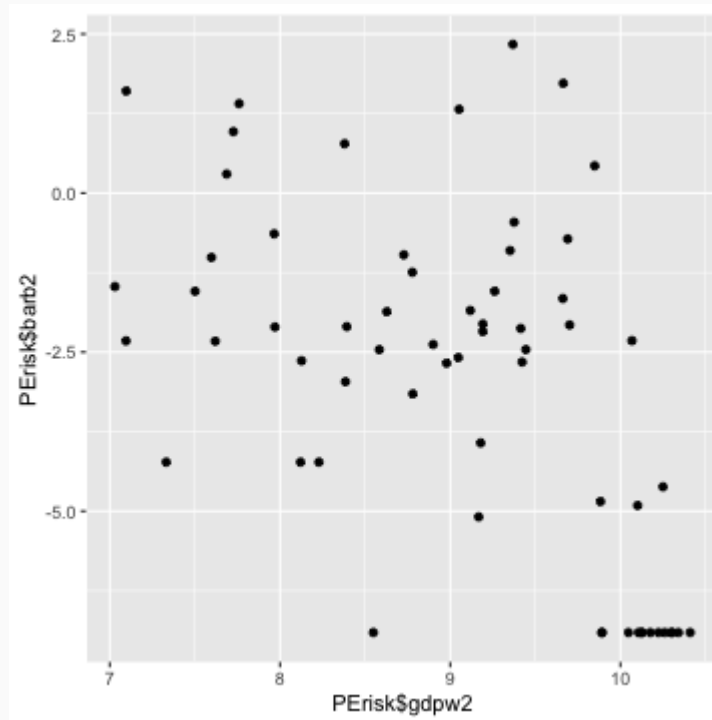


- **Your turn:** make a bar-plot of the `smoker` variable in the `tips` dataset.

Plot for two Continuous Variables

qplot - Scatter-Plot

```
qplot(PErisk$gdpw2, PErisk$barb2,  
      geom = "point")
```

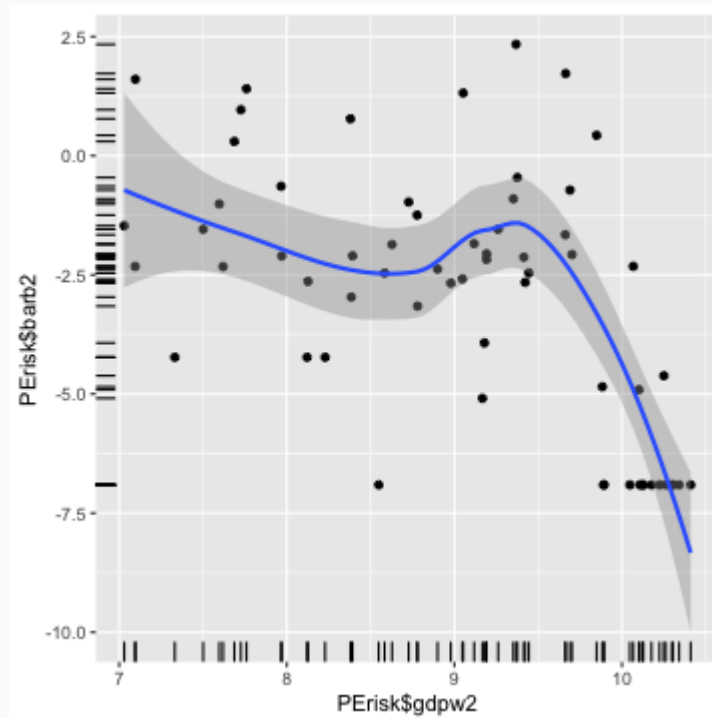


- **Your turn:** make a Scatter-Plot for the `totbill` and the `tips` variables in the `tips` dataset.

qplot - Scatter-Plot (with smooth)

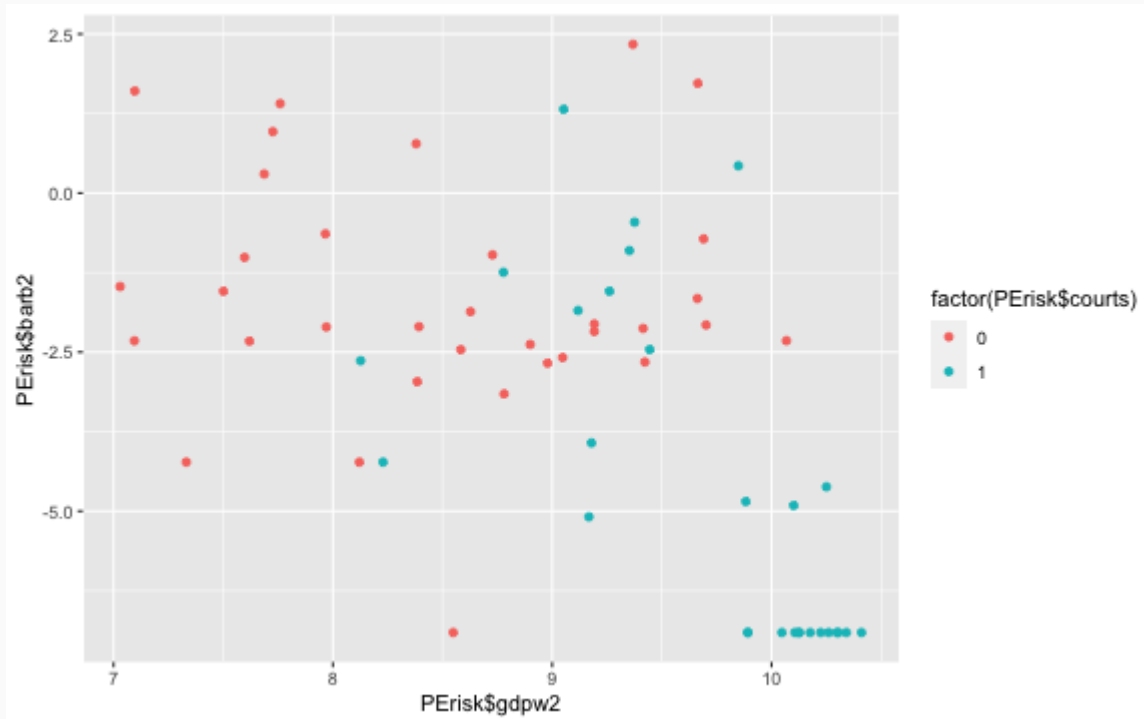
```
qplot(PErisk$gdpw2, PErisk$barb2, geom = "point") +  
  geom_rug() + geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



qplot - Scatter-Plot (segmented)

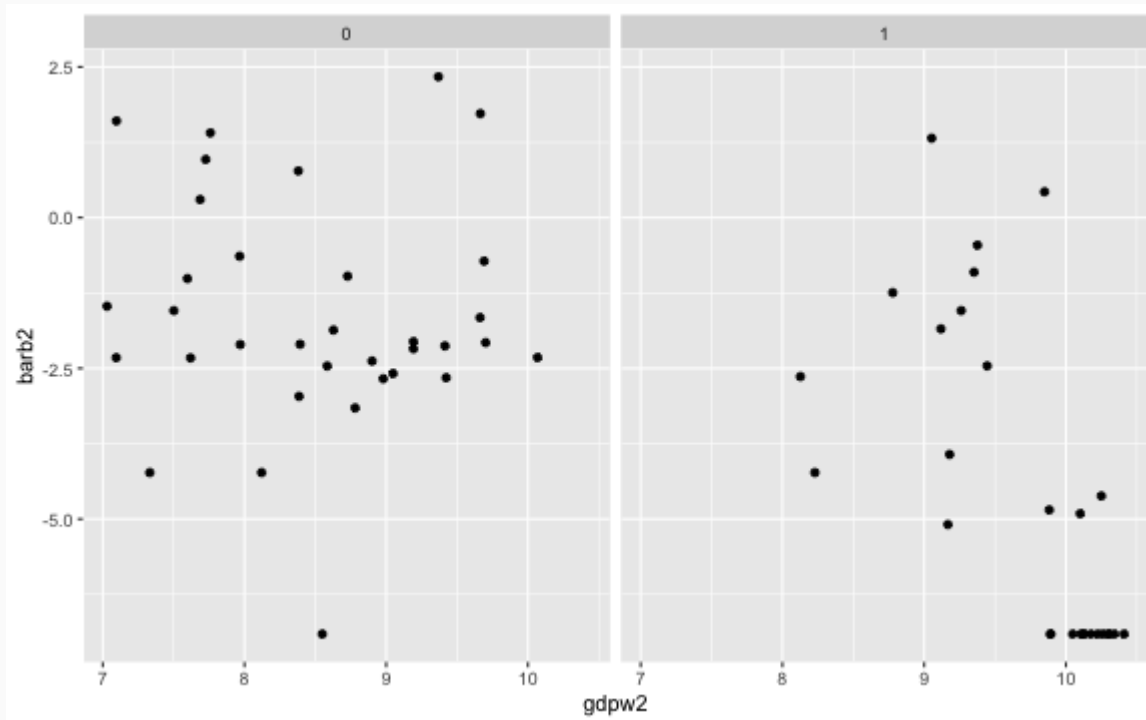
```
qplot(PERisk$gdpw2, PERisk$barb2, geom = "point", color = factor(PERisk$courts))
```



- **Your turn:** make a segmented scatter-plot with smooth line for the `totbill` and the `tips`, by `smoker`.

qplot - Scatter-Plot (faceted)

```
qplot(gdpw2, barb2, geom = "point", facets = . ~ courts, data = PE
```

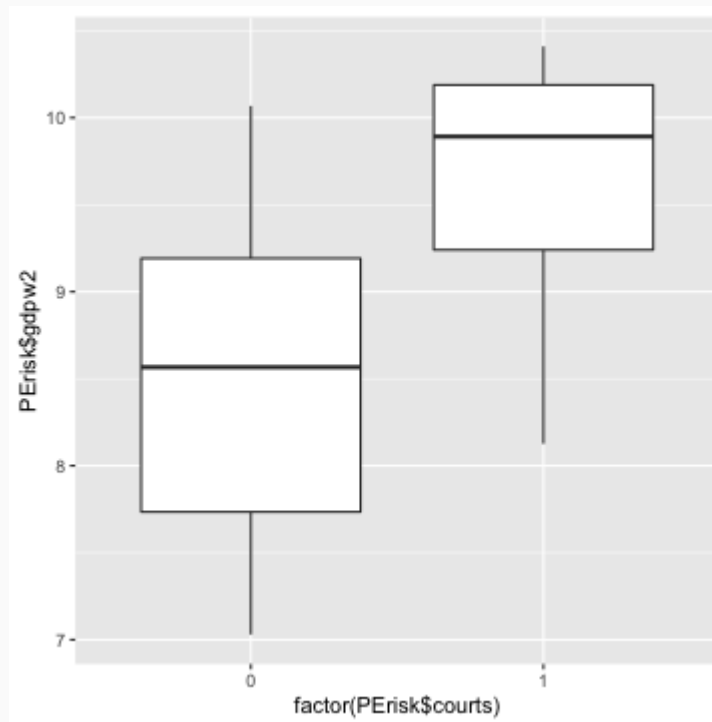


- **Your turn:** make a faceted scatter-plot for the `totbill` and the `tips`, faceting by `smoker`.

Plot for Continuous x Discrete Variables

qplot - Multiple Box-Plots

```
qplot(x = factor(PERisk$courts), y = PERisk$gdpw2,  
      geom = "boxplot")
```



- **Your turn:** make a box-plot of the `tips` variable by `smoker` in the `tips` dataset.

Questions?

Have a great weekend!
