QTM 150

Week 5 – Graphs

Umberto Mignozzetti
Feb 26

Recap

You now know:

- The main objects in R:
- How to do basic operations with datasets.

Do you have any questions?

- **Great job!!**

- **Reminder:** The quiz for this class will be posted at 4:00PM. The quiz is due Monday, 11:59 PM.

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 and ggplot

- `qplot`, as the names says, stands for quick plots.
- It is great to generate `ggplot` graphs in a hurry.
- `ggplot` graphs are the most basic and pretty graphs in R.
- `ggplot` implements what we call the *grammar of graphs*, which is a method to generate data viz that we are going to learn in this class.

qplot and ggplot

`ggplot` (and `qplot`) is based on the idea that graphs are a sum of three components:

- geoms (shapes)
- a dataset
- and a mapping system.

This is the **`grammar of graphs`**.

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.

Loading tidyverse

```
# Load tidyverse
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyv

## ✓ 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_c
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Loading datasets

```r
# Loading tips dataset
tips ← read.csv('https://raw.githubusercontent.com/umbertomig/qtm
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
```

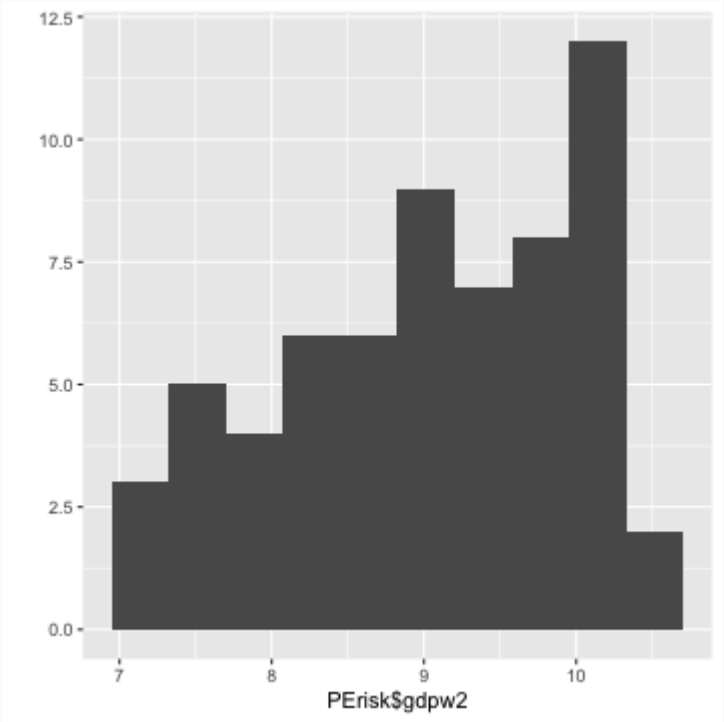
```r
# Loading PErisk dataset
PErisk ← read.csv('https://raw.githubusercontent.com/umbertomig/c
head(PErisk, 2)
```

```
##     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

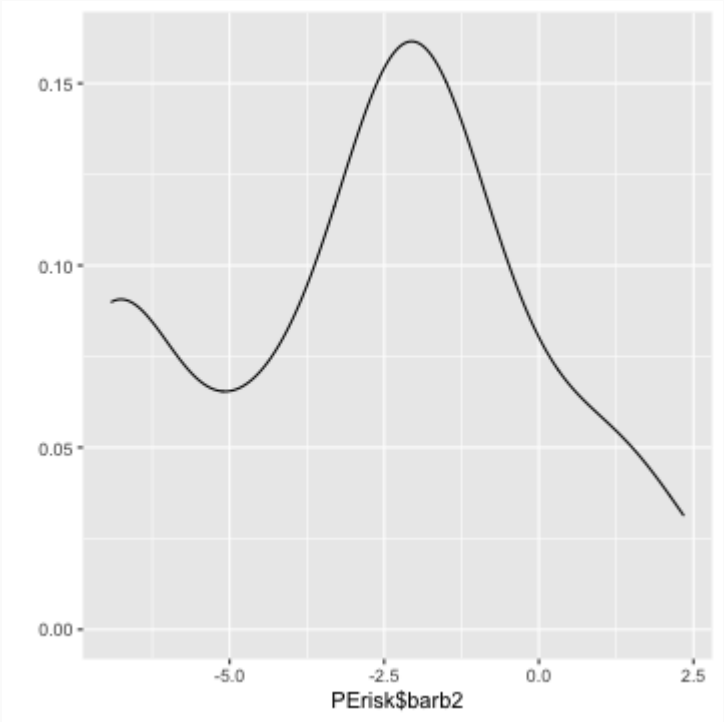
```
qplot(PErisk$gdpw2, geom = "histogram", bins = 10)
```



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

qplot - Density-plots

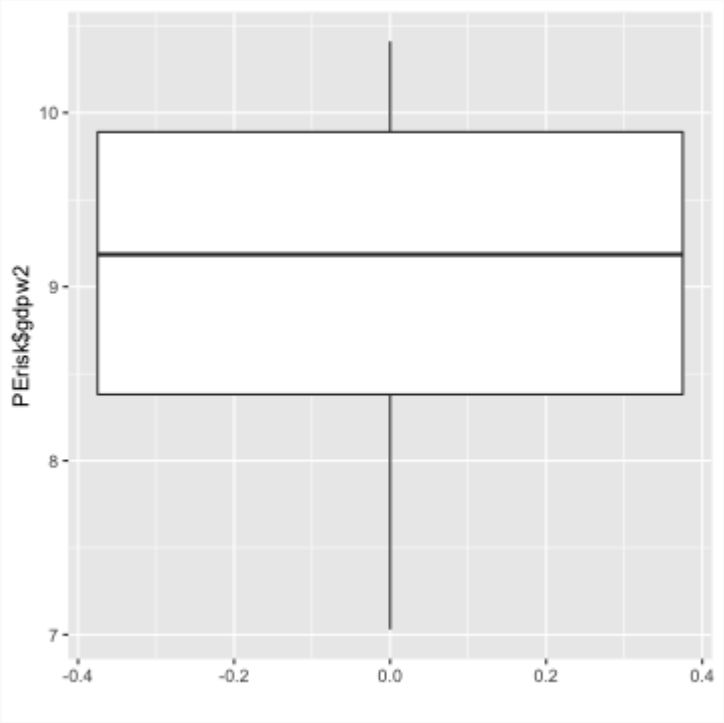
```
qplot(PErisk$barb2, geom = "density")
```



- **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")
```

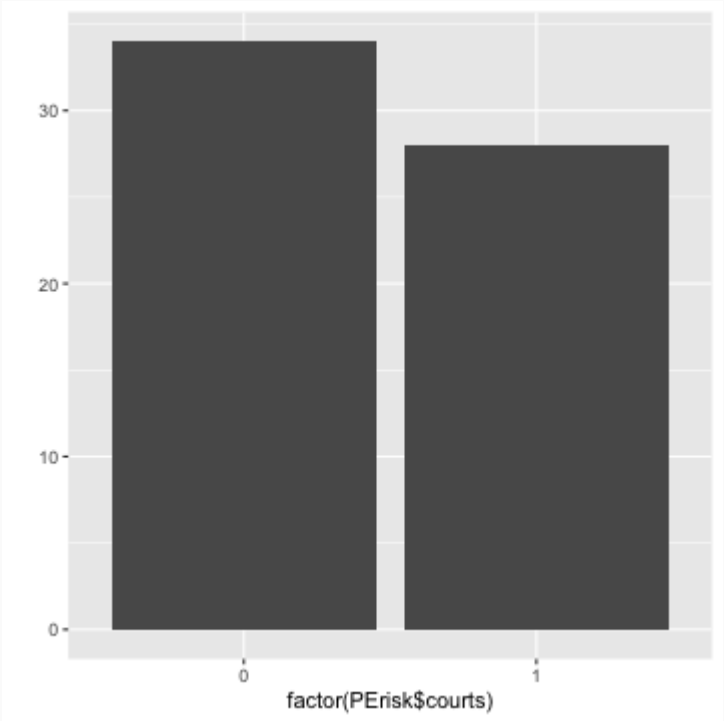


- **Your turn**: make a boxplot of the `totbill` variable in the `tips` dataset.

Plot for Discrete Variables

qplot - Bar-Plots

```
# Bar-plot of courts
qplot(factor(PErisk$courts), geom = "bar")
```

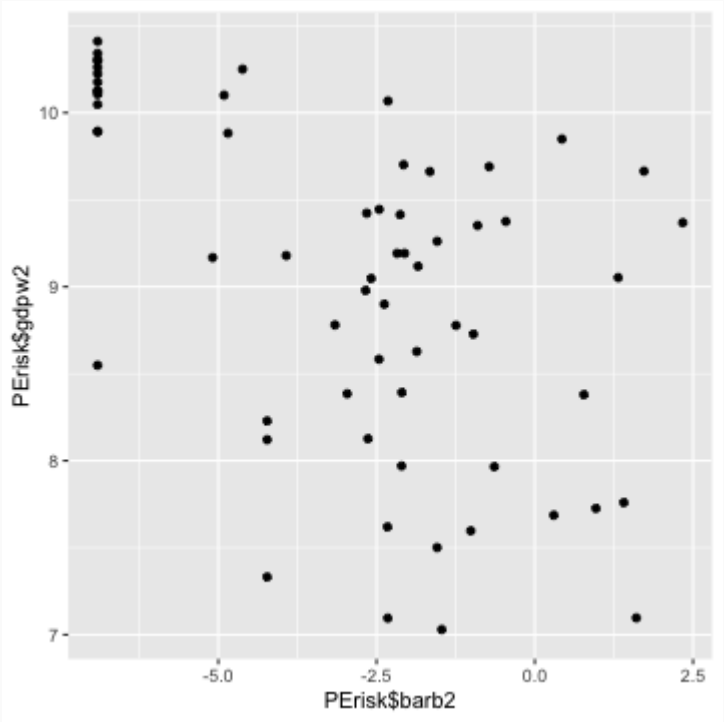


- **Your turn**: make a bar-plot of the `smoker` variable in the `tips` dataset.

Plot for two Continuous Variables

qplot - Scatter-Plot

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

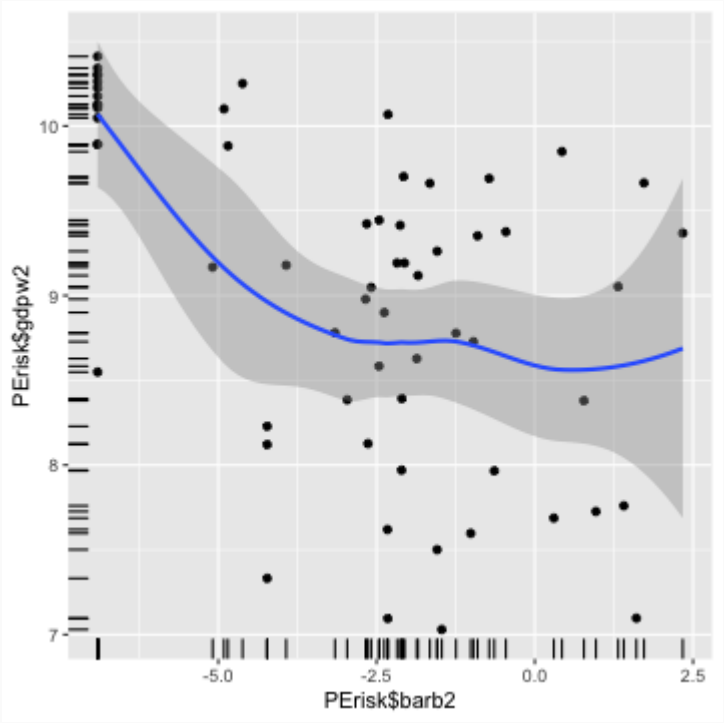


- **Your turn**: make a Scatter-Plot for the `totbill` and the `tips` variables in the `tips` dataset.

qplot - Scatter-Plot (with smooth)

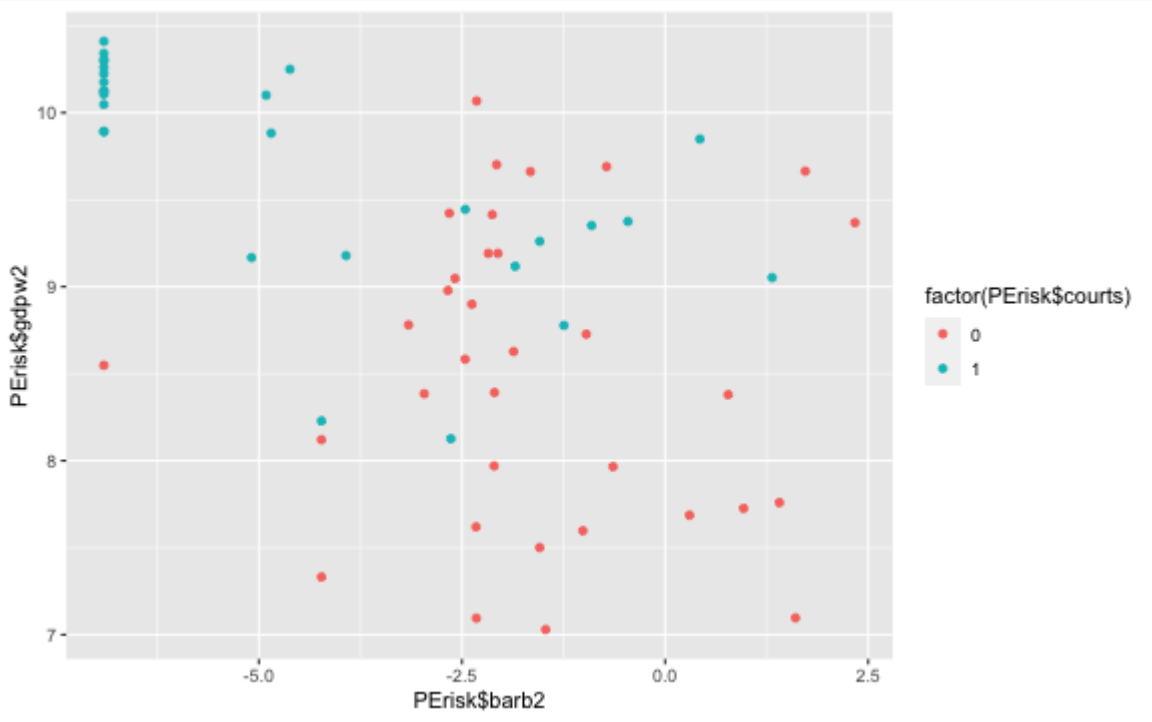
```
qplot(PErisk$barb2, PErisk$gdpw2, geom = "point") +
  geom_rug() + geom_smooth()
```

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



qplot - Scatter-Plot (segmented)

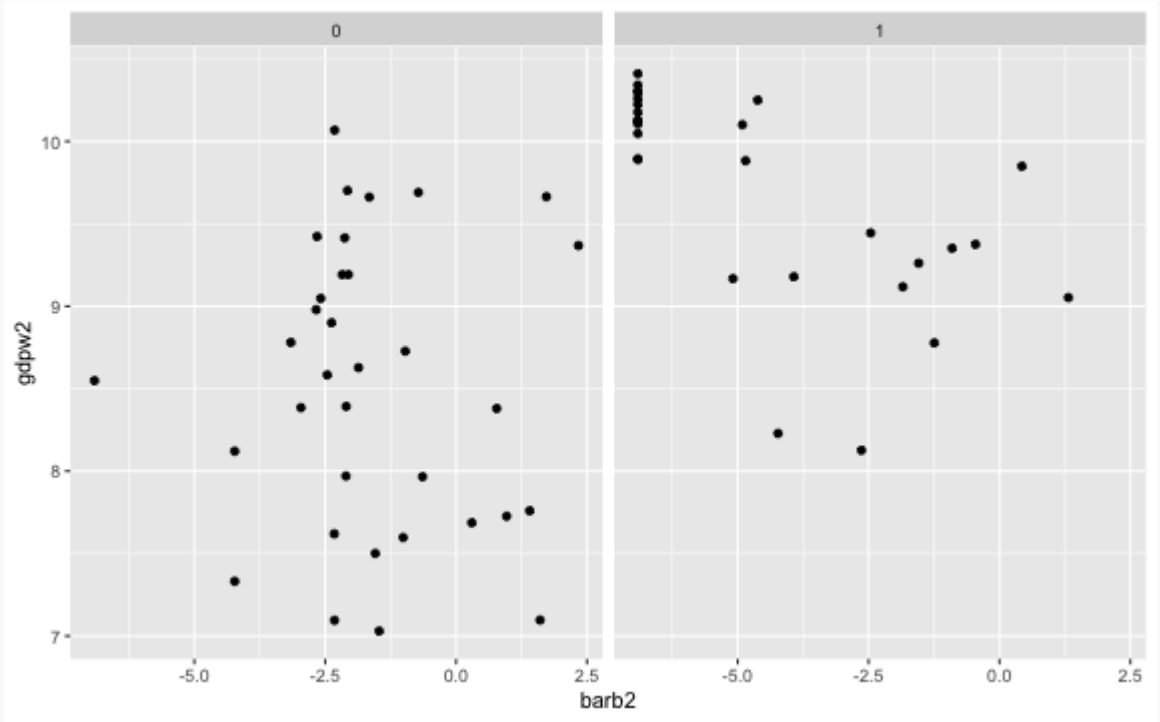
```
qplot(PErisk$barb2, PErisk$gdpw2, geom = "point", color = factor(P
```



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

qplot - Scatter-Plot (faceted)

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

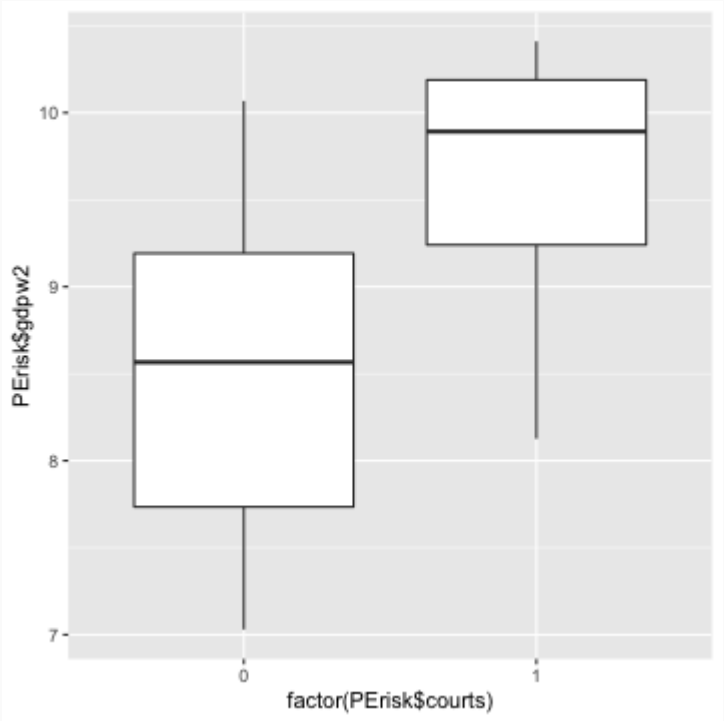


- **Your turn**: make a faceted scatter-plot for the `totbill` and the `tips`, faceting by `smoker`.

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!