Hexbin Chart and Contour 2D Plot

Thanh-Giang Tan Nguyen,
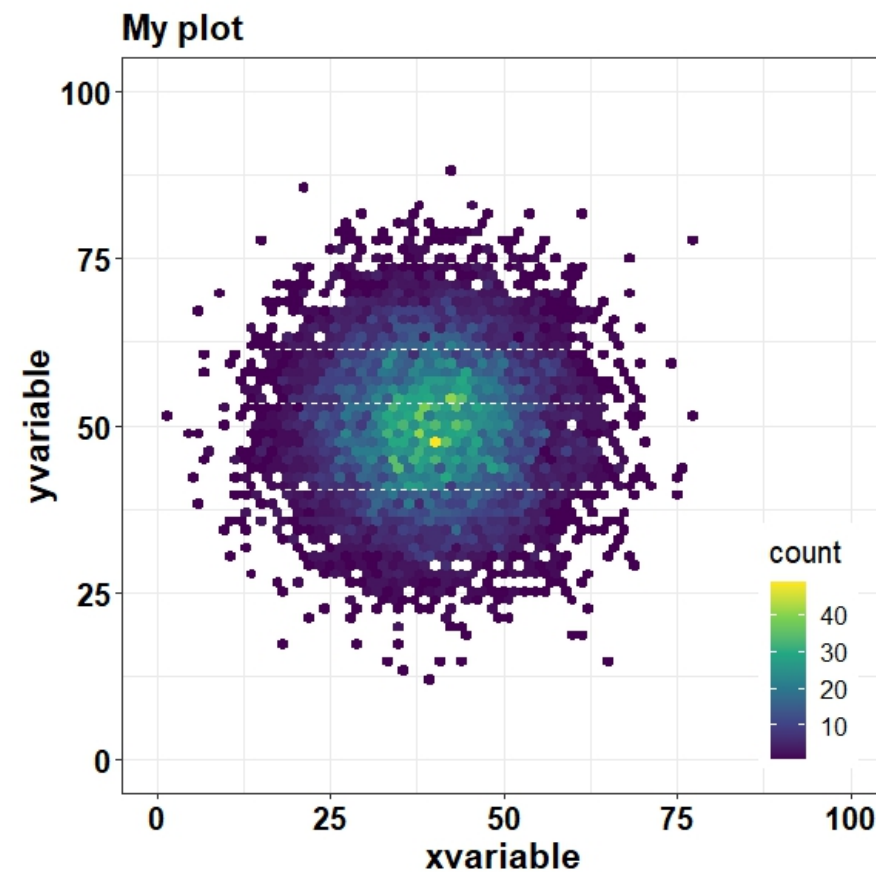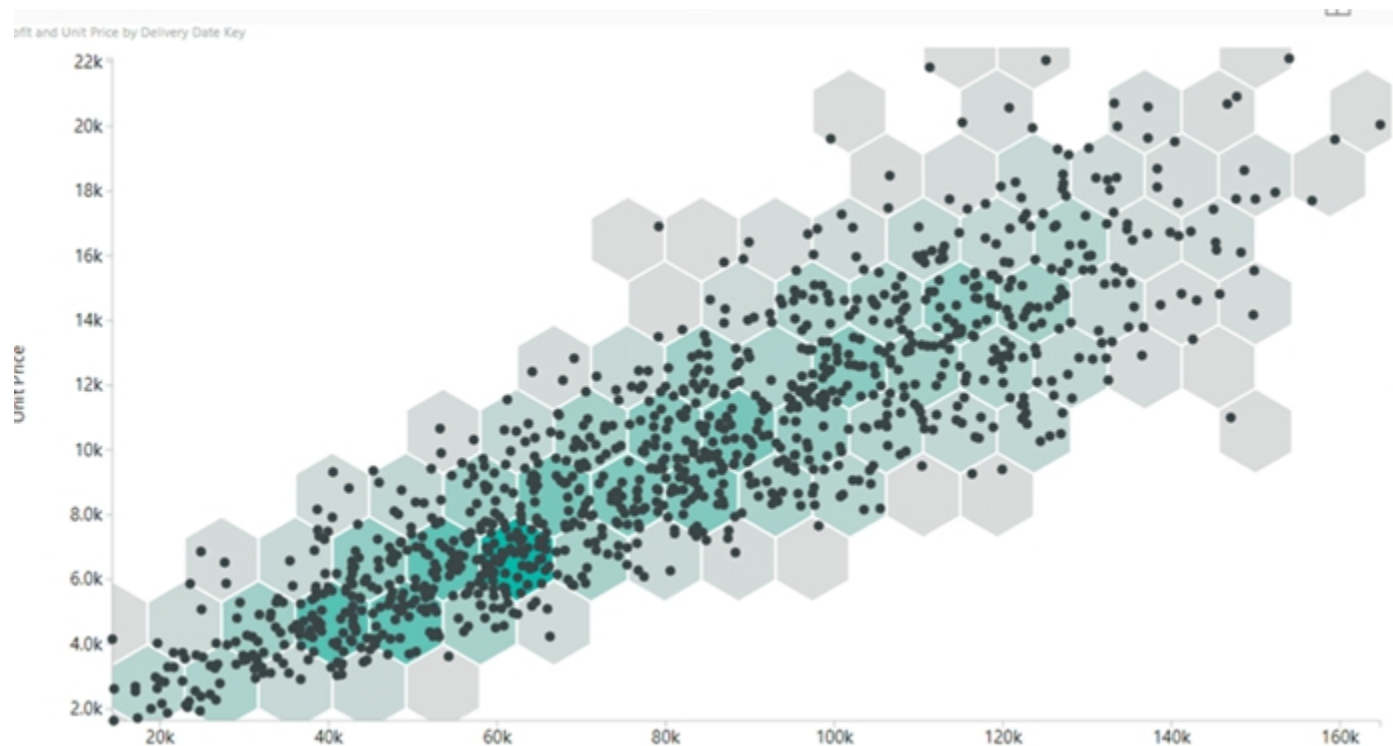ID:1805015
Course: AB-403
Data Visualization
Instructor: Phuc Loi Luu, PhD
p.luu@garvan.org.au
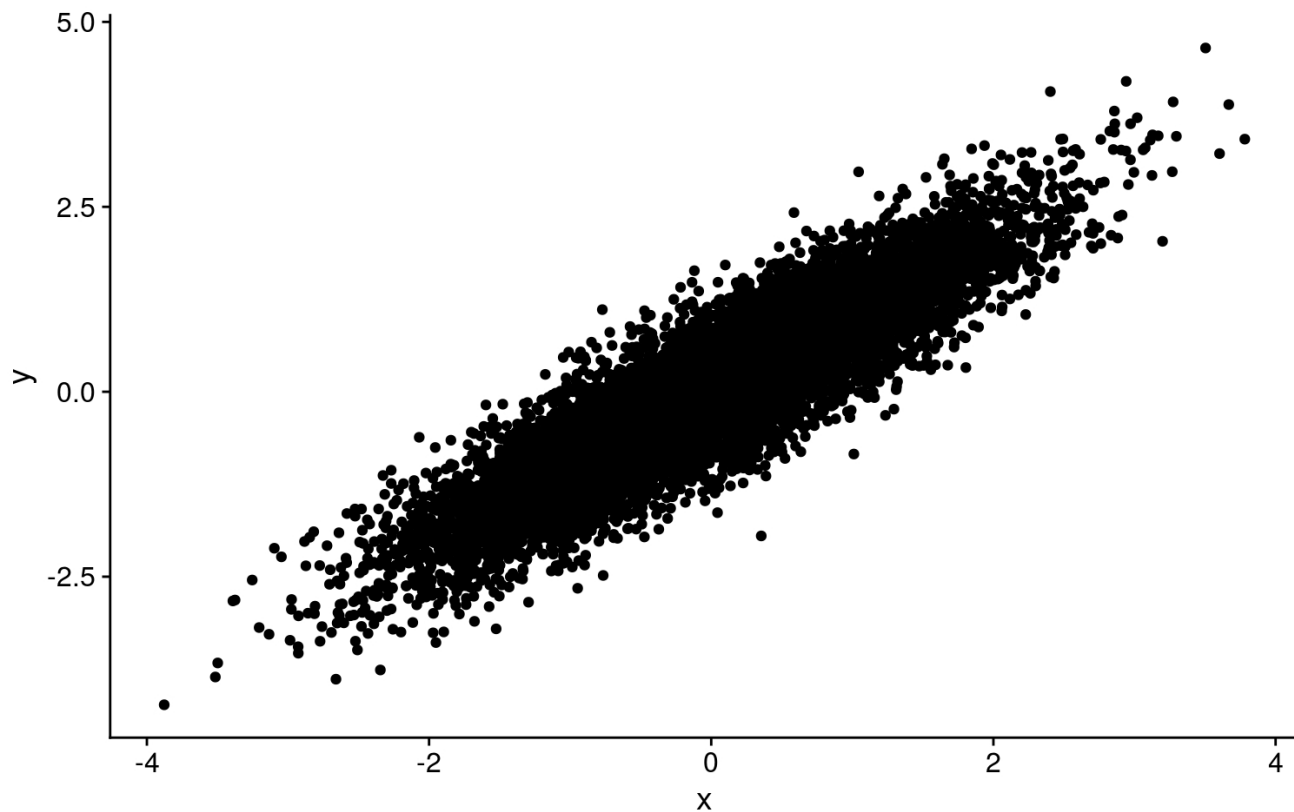luu.p.loi@googlemail.com

Adapted and modified from https://web.stanford.edu/~hastie/MOOCSlides/linear_regression.pdf

Hexbin Chart

- Instead of using the scatter plot, the dots may be overlaped, therefore, using hexagonal plot is much better.
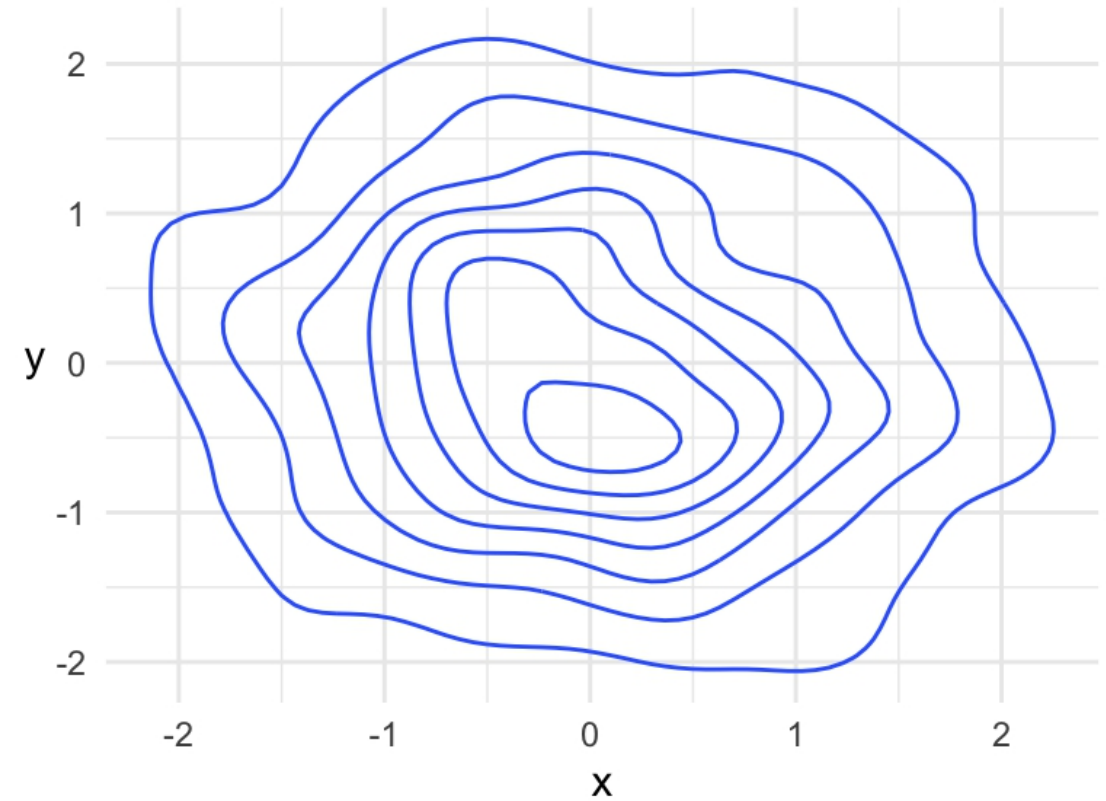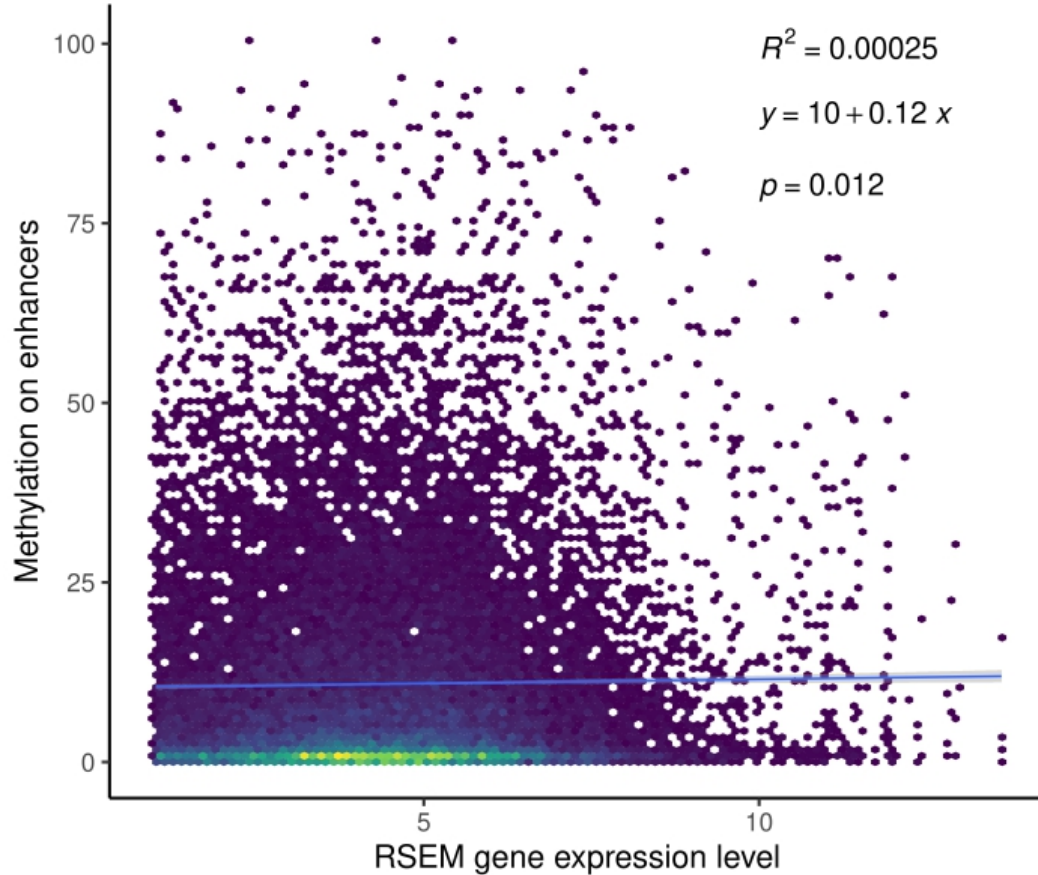- The binning size and color bar should be considered for a better visualization.

Hexbin Chart: Practice

$R^2 = 0.00025$

$y = 10 + 0.12\,x$

$p = 0.012$

- Using the simulated data provided by script, plot the hexagonal plot as the right figure.

Contour Plot

$R^2 = 0.00025$

$y = 10 + 0.12\,x$

$p = 0.012$
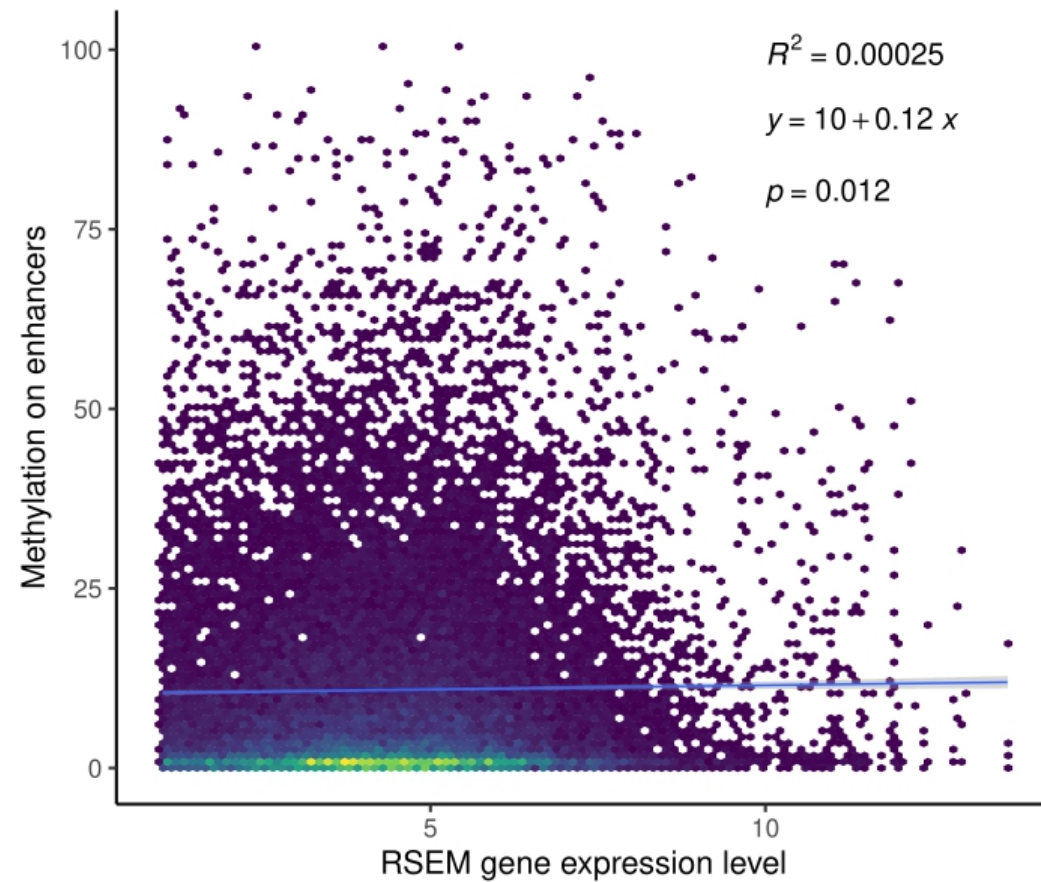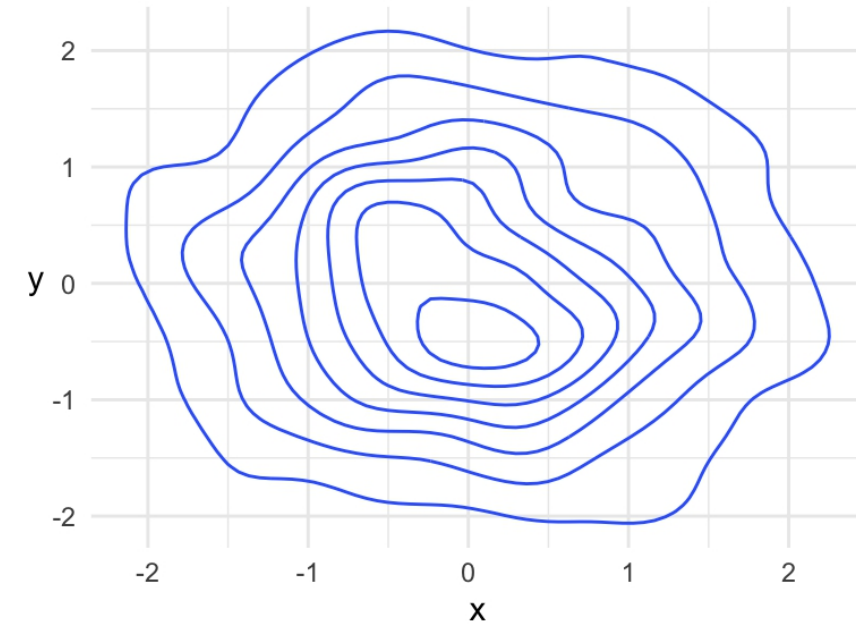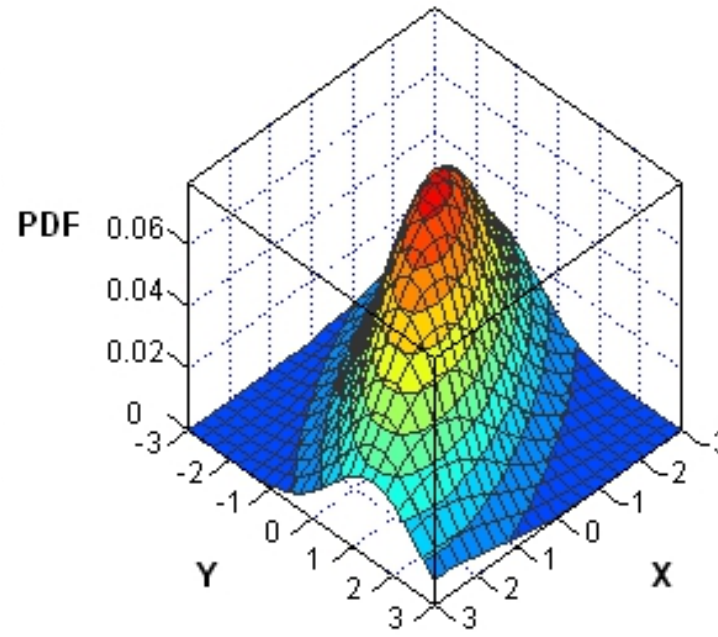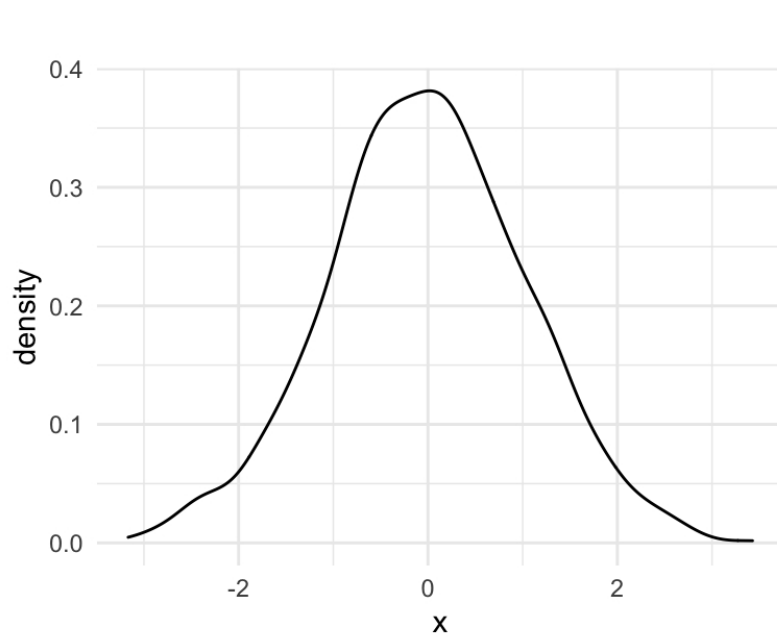
- The contour plot is quite similar to hexbin plot, the line will cover the data inside with the ralative colors. However, it is based on the distribution of data

Contour Plot

- Density plot is based on the data, give you an plot for estimating the distribution of data. The shape of density plot can be estimated using histogram with the high number of bins.
- Using the second dimension (y), bivariate kernel density estimate can be draw similarly. The contour will be drawed from the fundamentel elements in figure A.

Kernel Density Plot: https://www.youtube.com/watch?v=x5zLaWT5KPs
https://www.data-to-viz.com/graph/density2d.html

Contour Plot: Practice
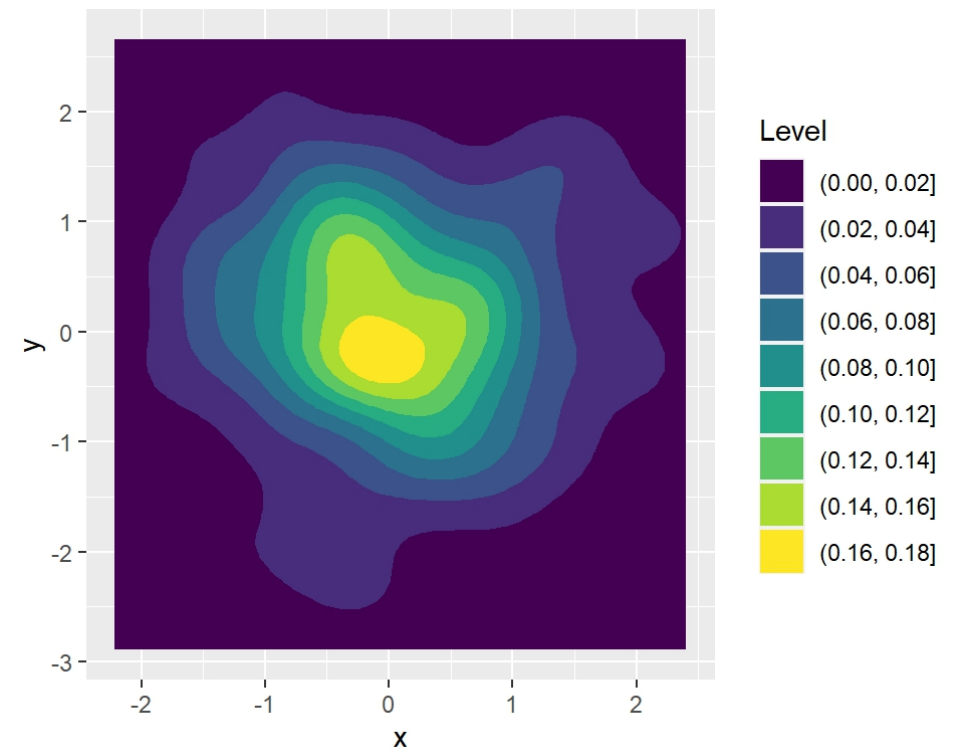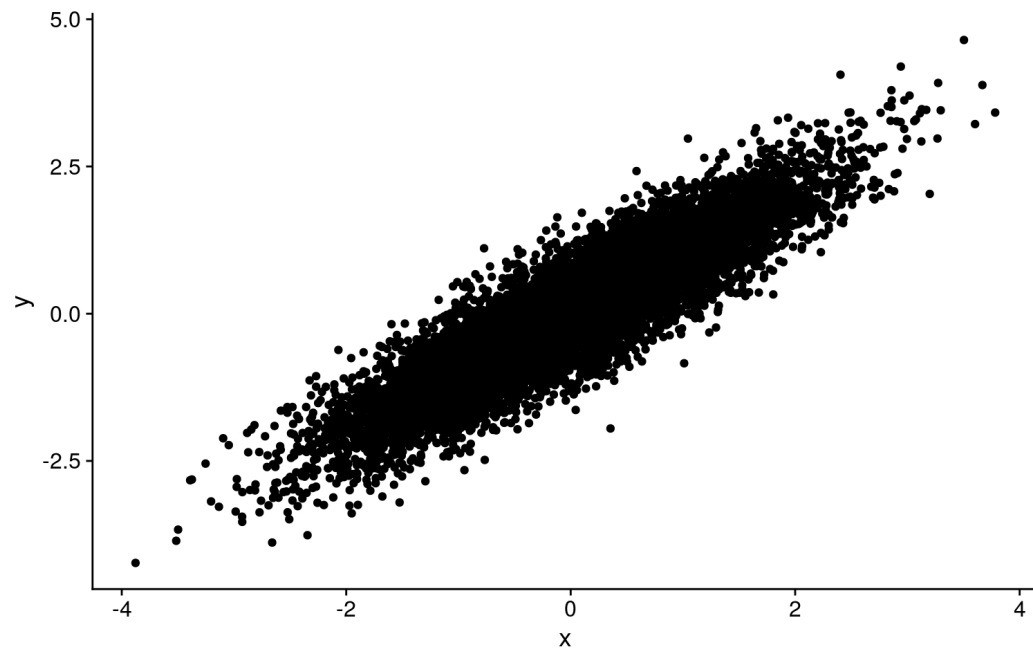
- Using the data in previous section, draw the similar contour plot on the right figure with bin size=20

Tryhard: Using the methyltion 27k TCGA BRCA to visualize:

01. On one people code: TCGA-BH-A0B7 with 2 respective samples for normal and cancer:

TCGA-BH-A0B7-01A: Cancer sample

TCGA-BH-A0B7-11A: Normal sample

Visualize the correlation between normal and cancer DNA methylation on each CpG site using 27k array data.

Sample	Sample type	1	A solid tumor	Tumor types range from 01 - 09, normal types from 10 - 19 and control samples from 20 - 29. See Code Tables Report for a complete list of sample codes

Tryhard: Using the methyltion 27k TCGA BRCA to visualize:

01. On one people code: TCGA-BH-A0B7 with 2 respective samples for normal and cancer:

TCGA-BH-A0B7-01A: Cancer sample

TCGA-BH-A0B7-11A: Normal sample

Visualize the correlation between normal and cancer DNA methylation on each CpG site using 27k array data.

Hint:

```
> meth
          probe          variable       value          type
   1: cg00000292 TCGA-A2-A0EQ-01A 0.68929498 TCGA-A2-A0EQ
   2: cg00002426 TCGA-A2-A0EQ-01A 0.55042375 TCGA-A2-A0EQ
   3: cg00003994 TCGA-A2-A0EQ-01A 0.38996204 TCGA-A2-A0EQ
   4: cg00005847 TCGA-A2-A0EQ-01A 0.53378015 TCGA-A2-A0EQ
   5: cg00006414 TCGA-A2-A0EQ-01A         NA TCGA-A2-A0EQ
  ---
9431672: cg27657283 TCGA-BH-A18L-11A 0.05898135 TCGA-BH-A18L
9431673: cg27661264 TCGA-BH-A18L-11A 0.25390677 TCGA-BH-A18L
9431674: cg27662379 TCGA-BH-A18L-11A 0.01036165 TCGA-BH-A18L
9431675: cg27662877 TCGA-BH-A18L-11A 0.05782251 TCGA-BH-A18L
9431676: cg27665659 TCGA-BH-A18L-11A 0.02643607 TCGA-BH-A18L
```

```
> cancer
          probe          variable       value          type
   1: cg00000292 TCGA-A2-A0EQ-01A 0.68929498 TCGA-A2-A0EQ
   2: cg00002426 TCGA-A2-A0EQ-01A 0.55042375 TCGA-A2-A0EQ
   3: cg00003994 TCGA-A2-A0EQ-01A 0.38996204 TCGA-A2-A0EQ
   4: cg00005847 TCGA-A2-A0EQ-01A 0.53378015 TCGA-A2-A0EQ
   5: cg00007981 TCGA-A2-A0EQ-01A 0.03011270 TCGA-A2-A0EQ
  ---
7714560: cg27657249 TCGA-C8-A134-01A 0.32844675 TCGA-C8-A134
7714561: cg27661264 TCGA-C8-A134-01A 0.44077153 TCGA-C8-A134
7714562: cg27662379 TCGA-C8-A134-01A 0.50920786 TCGA-C8-A134
7714563: cg27662877 TCGA-C8-A134-01A 0.06509386 TCGA-C8-A134
7714564: cg27665659 TCGA-C8-A134-01A 0.01455412 TCGA-C8-A134
```

```
> normal
          probe          variable       value          type
   1: cg00000292 TCGA-E2-A15M-11A 0.60197442 TCGA-E2-A15M
   2: cg00002426 TCGA-E2-A15M-11A 0.24883948 TCGA-E2-A15M
   3: cg00003994 TCGA-E2-A15M-11A 0.05485671 TCGA-E2-A15M
   4: cg00005847 TCGA-E2-A15M-11A 0.52714029 TCGA-E2-A15M
   5: cg00006414 TCGA-E2-A15M-11A         NA TCGA-E2-A15M
  ---
744602: cg27657283 TCGA-BH-A18L-11A 0.05898135 TCGA-BH-A18L
744603: cg27661264 TCGA-BH-A18L-11A 0.25390677 TCGA-BH-A18L
744604: cg27662379 TCGA-BH-A18L-11A 0.01036165 TCGA-BH-A18L
744605: cg27662877 TCGA-BH-A18L-11A 0.05782251 TCGA-BH-A18L
744606: cg27665659 TCGA-BH-A18L-11A 0.02643607 TCGA-BH-A18L
```

```
> df
          probe       type     variable.x    value.x     variable.y    value.y
   1: cg00000292 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.42539105 TCGA-BH-A0B7-11A 0.70821517
   2: cg00000292 TCGA-BH-A0BL TCGA-BH-A0BL-01A 0.66201871 TCGA-BH-A0BL-11A 0.63411007
   3: cg00000292 TCGA-BH-A0BO TCGA-BH-A0BO-01A 0.58211707 TCGA-BH-A0BO-11A 0.61125256
   4: cg00000292 TCGA-BH-A0BQ TCGA-BH-A0BQ-01A 0.62525161 TCGA-BH-A0BQ-11A 0.64348443
   5: cg00000292 TCGA-BH-A0BW TCGA-BH-A0BW-01A 0.42091990 TCGA-BH-A0BW-11A 0.74593300
  ---
648838: cg27665659 TCGA-E2-A14Y TCGA-E2-A14Y-01A 0.01176157 TCGA-E2-A14Y-11A 0.03133620
648839: cg27665659 TCGA-E2-A153 TCGA-E2-A153-01A 0.03972280 TCGA-E2-A153-11A 0.01935532
648840: cg27665659 TCGA-E2-A158 TCGA-E2-A158-01A 0.01443324 TCGA-E2-A158-11A 0.04592540
648841: cg27665659 TCGA-E2-A15L TCGA-E2-A15L-01A 0.01845402 TCGA-E2-A15L-11A 0.03635290
648842: cg27665659 TCGA-E2-A15M TCGA-E2-A15M-01A 0.02387033 TCGA-E2-A15M-11A 0.03363276
```

```
> df1
          probe       type     variable.x     value.x      variable.y     value.y
   1: cg00000292 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.425391048 TCGA-BH-A0B7-11A 0.708215171
   2: cg00002426 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.333198595 TCGA-BH-A0B7-11A 0.062914963
   3: cg00003994 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.077630224 TCGA-BH-A0B7-11A 0.031415520
   4: cg00005847 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.515929674 TCGA-BH-A0B7-11A 0.459821932
   5: cg00007981 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.060211190 TCGA-BH-A0B7-11A 0.048131805
  ---
24974: cg27657283 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.053344254 TCGA-BH-A0B7-11A 0.039038029
24975: cg27661264 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.369958746 TCGA-BH-A0B7-11A 0.220298966
24976: cg27662379 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.009150523 TCGA-BH-A0B7-11A 0.009478352
24977: cg27662877 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.023082475 TCGA-BH-A0B7-11A 0.032744624
24978: cg27665659 TCGA-BH-A0B7 TCGA-BH-A0B7-01A 0.012001582 TCGA-BH-A0B7-11A 0.013391843
>
```
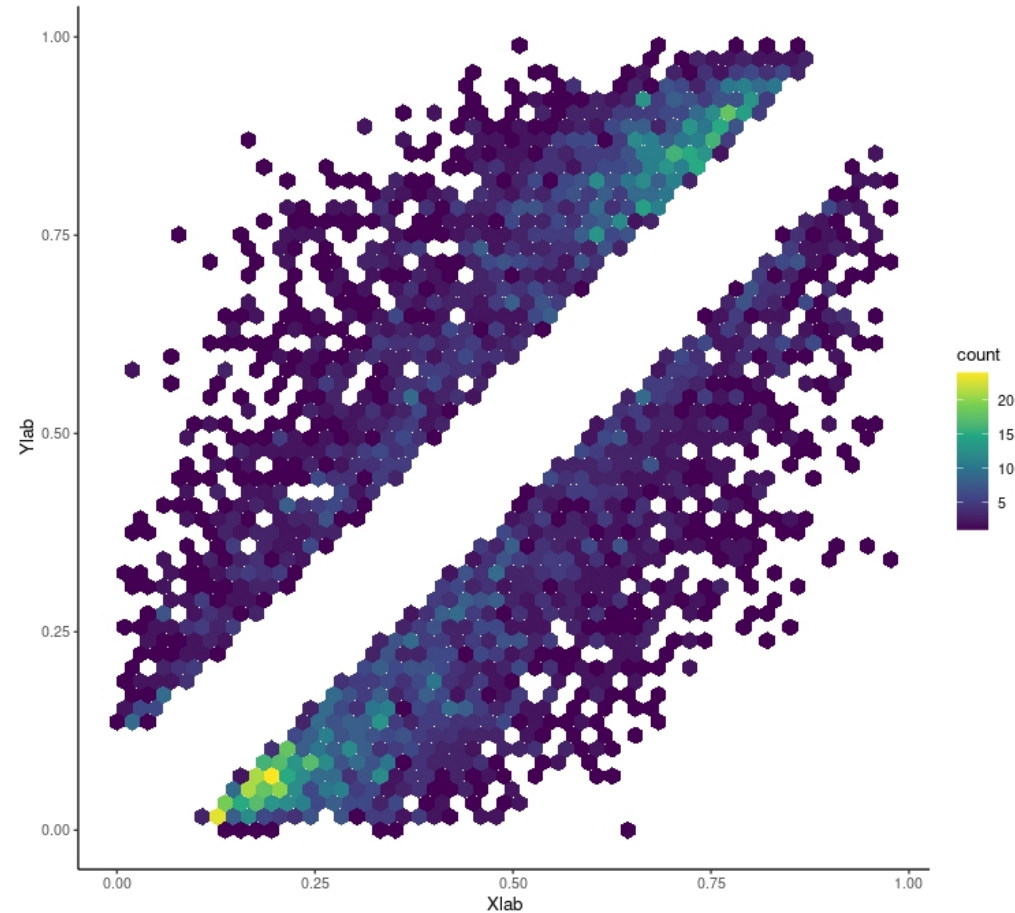
Tryhard: Using the methyltion 27k TCGA BRCA to visualize:

02. Using the dataframe for visualization in 01, remove the datapoint where the differential level between normal and cancer smaller than 0.1 (10%)
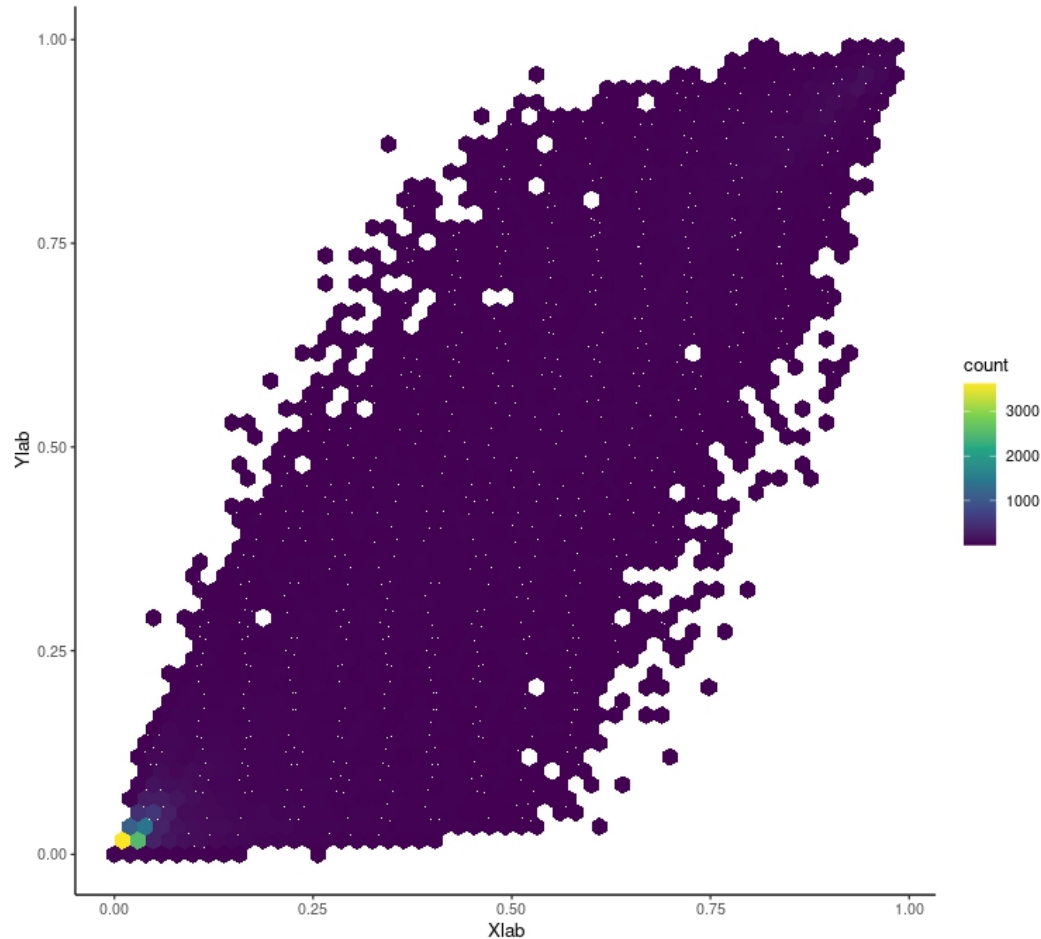
Tryhard: Using the methyltion 27k TCGA BRCA to visualize:

03. Similar to 01, but group by CpG sites, calculate the mean of each groups (normal or cancer), then visualize it.


```
> cancer
          probe       mean
   1: cg00000292 0.57869002
   2: cg00002426 0.23730505
   3: cg00003994 0.15839444
   4: cg00005847 0.62610673
   5: cg00007981 0.05024759
  ---
24977: cg27657283 0.03969442
24978: cg27661264 0.35351354
24979: cg27662379 0.01305073
24980: cg27662877 0.03830513
24981: cg27665659 0.02360740
>
```

```
> normal
          probe       mean
   1: cg00000292 0.61018022
   2: cg00002426 0.14790153
   3: cg00003994 0.04606054
   4: cg00005847 0.52822655
   5: cg00007981 0.03828241
  ---
24977: cg01288598 0.10856009
24978: cg16199381 0.02761262
24979: cg22385764 0.03109319
24980: cg24642468 0.60555006
24981: cg24647726 0.48107839
>
```

```
> df2
          probe    mean.x     mean.y  abs_diff
   1: cg00003994 0.1583944 0.04606054 0.1123339
   2: cg00013618 0.7337084 0.85938157 0.1256732
   3: cg00015770 0.3183240 0.08277856 0.2355455
   4: cg00019495 0.1453294 0.36681461 0.2214852
   5: cg00027083 0.3061116 0.08158837 0.2245232
  ---
5415: cg27635271 0.4865318 0.27901207 0.2075197
5416: cg27637521 0.2281410 0.09816766 0.1299733
5417: cg27641018 0.4922183 0.72541484 0.2331965
5418: cg27652350 0.6324720 0.52753693 0.1049351
5419: cg27661264 0.3535135 0.20705315 0.1464604
```

Tryhard: Using the methyltion 27k TCGA BRCA to visualize:

04. Similar to 02, using the result from 03 filtering with 10% differences.