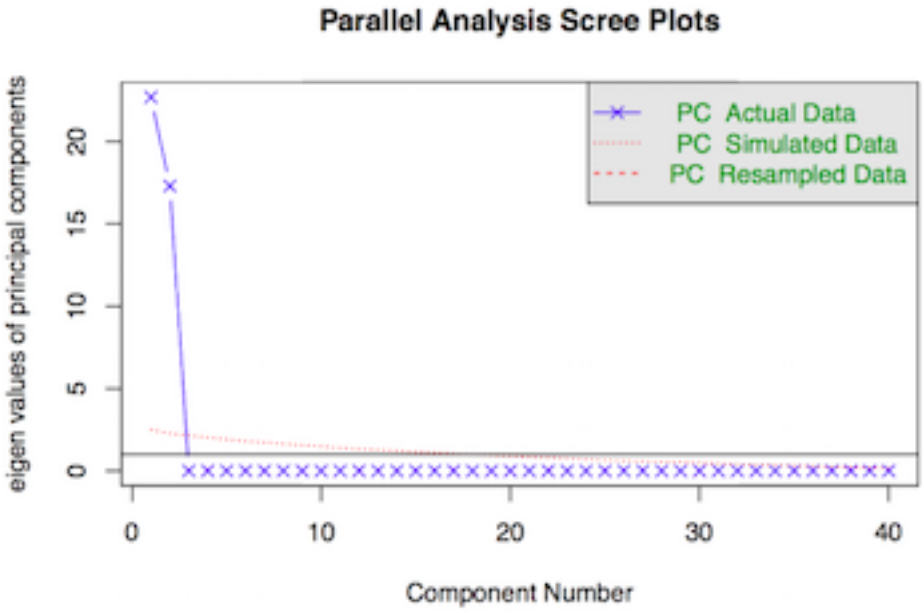
Lecture 8a

What should you be aware of when using the rainbow colormap with plots?

- All of these
- After the ROYGBIV colors are exhausted the colors will start repeating
- Certain colors, such as yellow, are hard to distinguish on some screens
- Our eyes tend to focus on colors that "pop", potentially biasing our interpretation of the plot

Lecture 8b

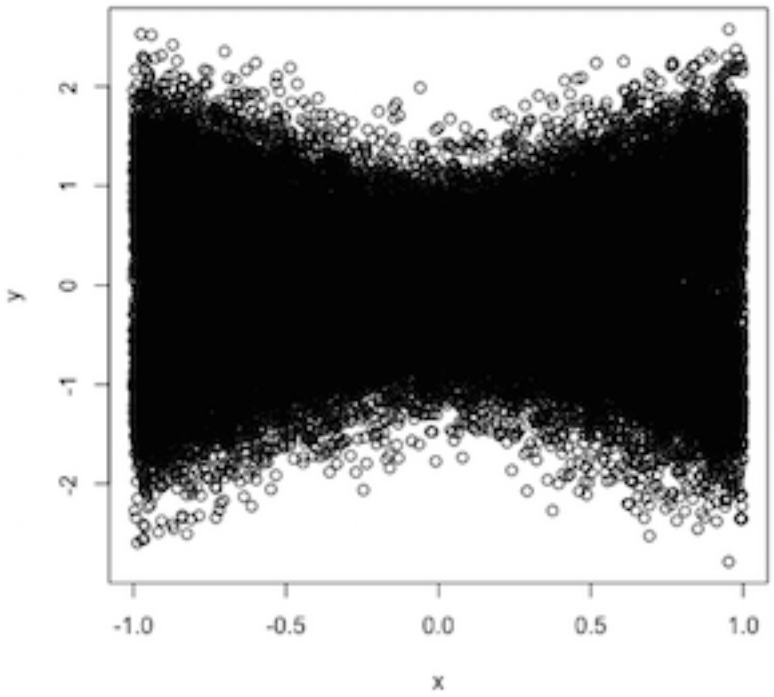
A scree plot displays eigenvalues associated with components or factors and can help determine the number of factors that display most of the varaiability in a given data set. Using the scree plot below, what is the ideal number of factors for PCA from this data set?



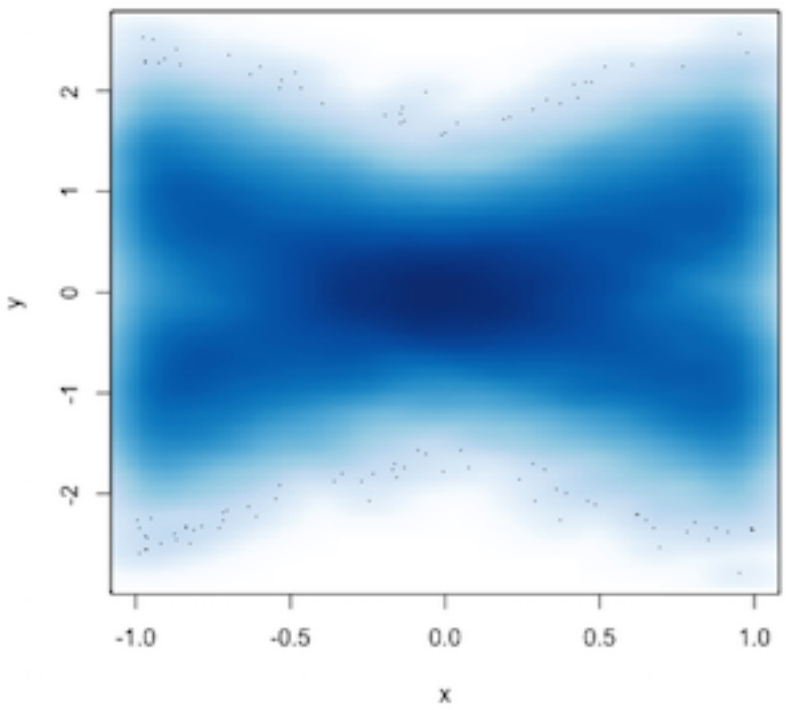
- 2
- 1
- 3
- 40

Lecture 9a

You create the plot below.



Realizing this display format does not tell you much about the data, you plot the same data, this time creating the new plot below.



What can be said about this new plot?

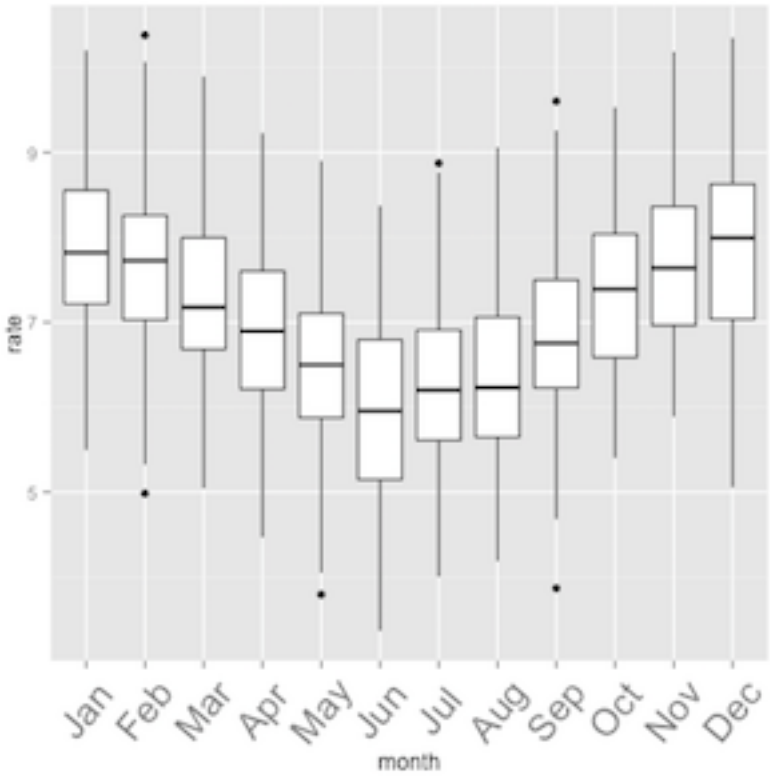
- It is a smooth scatter, or density plot, useful when a data set displays overplotting

- It is a blurred scatter, useful when a data set displays overplotting
- It not considered a scatter or density plot at all
- It requires a special package in R to produce

Lecture 9b

The following code generates the plot below:

```
g2 + geom_boxplot() + theme(axis.text.x=element_text(angle=50, size=20, vjust=0.5))
```



What do the box and whisker represent in the plot?

- The interquartile range (the middle half of the data set), and the median of the data, respectively
- The most common values in the data set, and the median of the data, respectively
- The interquartile range (the middle half of the data set), and the mean of the data, respectively
- The middle 25% of the data, and the median of the data, respectively

Lecture 10a

Shiny apps for R are good for:

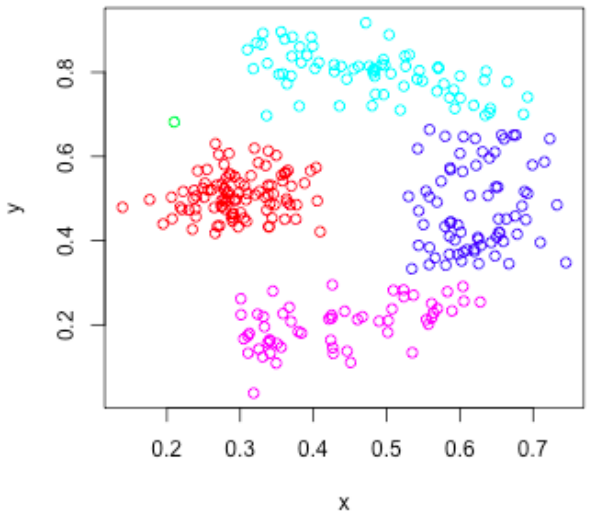
- All of these
- Exploratory data analysis
- Embedding interactions into R Presentations
- Sharing data sets with others on the web

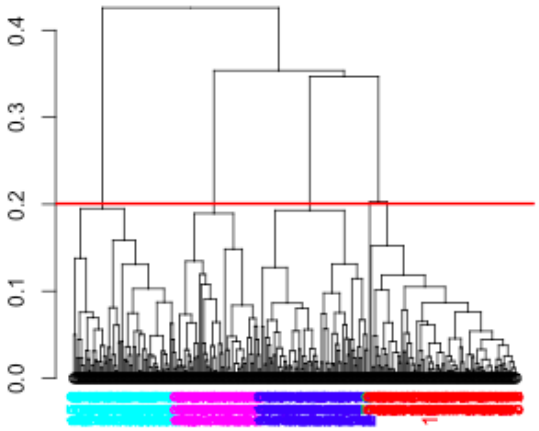
Lecture 10b

Shiny apps have a number of widgets available to change parameters in functions and plots. When adjusting, for example, the Slider Input while running a Shiny app, the rest of the data will update...

- immediately
- only after you refresh the page
- only if the developer updates the app
- never

Lecture 11a





The type of clustering shown above is called...

- hierarchical clustering
- density clustering
- K-means clustering
- relationship clustering

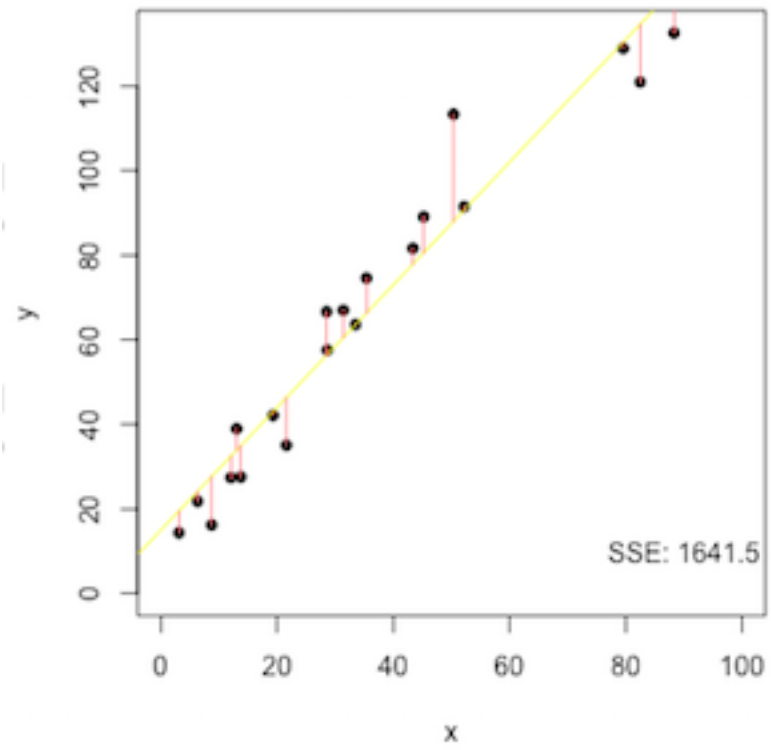
Lecture 11b

K-means clustering attempts to minimize...

- the within sum of squares value of each cluster
- the between sum of squares value
- the number of points in each cluster
- the randomness of cluster assignment

Lecture 12a

Linear regression attempts to minimize the sum of squares error between the vertical distances of points and the fitted line. What is the plot below showing?



- How the error between the fitted line and points is measured
- A random quadratic equation plotting against some random points
- A poorly fit linear equation
- None of these answers is correct

Lecture 12b

What is the name of a common function in R that fits linear models?

- `lm()`
- `lin()`
- 'linfit()‘
- `fit()`

Lecture 13a

Linear regression treats the input variables as...

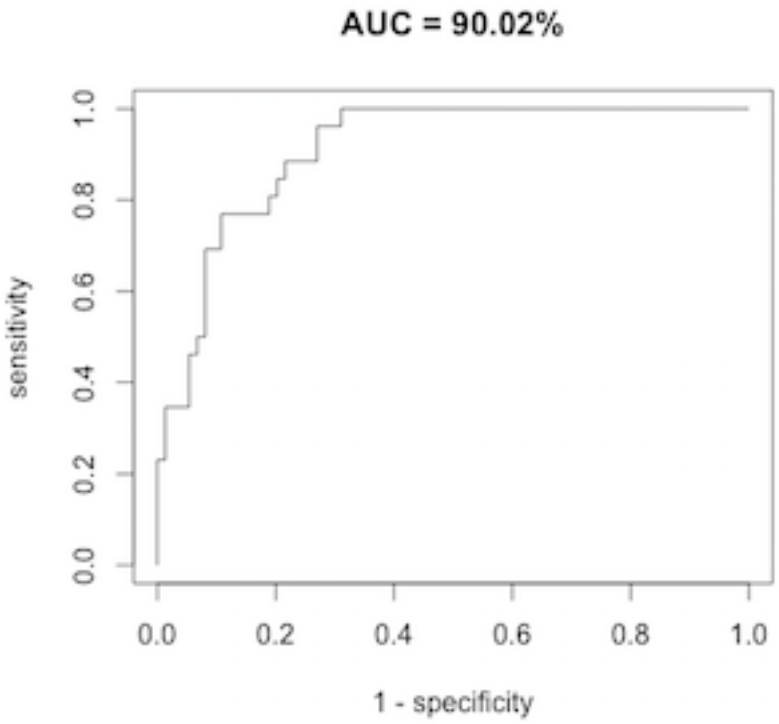
- numerical values
- boolean
- categories
- continuous densities

Lecture 13b

Which function included in R can be used to perform logistic regression?

- `glm()`
- `lm()`
- `fit()`
- `logit()`

Lecture 14a

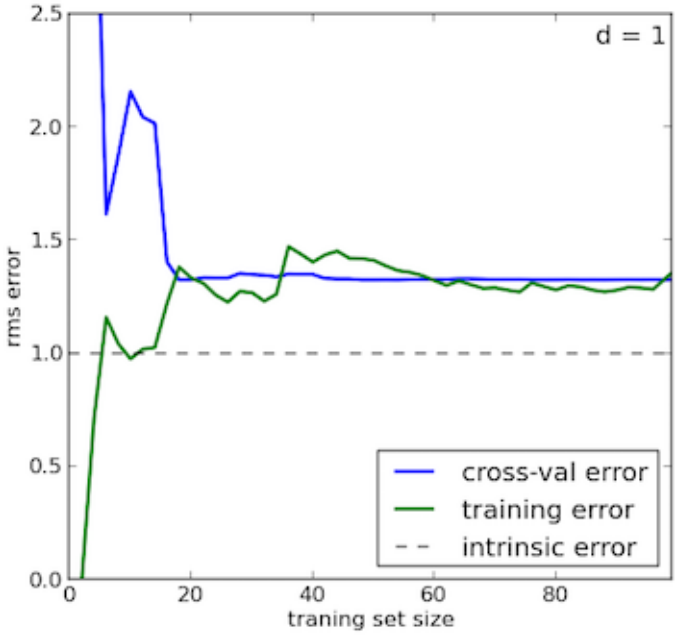


ROC curves, such as the one shown above, are a common tool used to...

- evaluate the performance of machine learning models
- evaluate the fit of a linear model
- plot points of a data set against predictions
- strike fear into the hearts of men

Lecture 14b

What can be said about the following learning curve?



- Adding additional points to the training data *will not* help improve the model
- Adding additional points to the training data *will* help improve the model
- The model is over-fitted
- The model is under-fitted