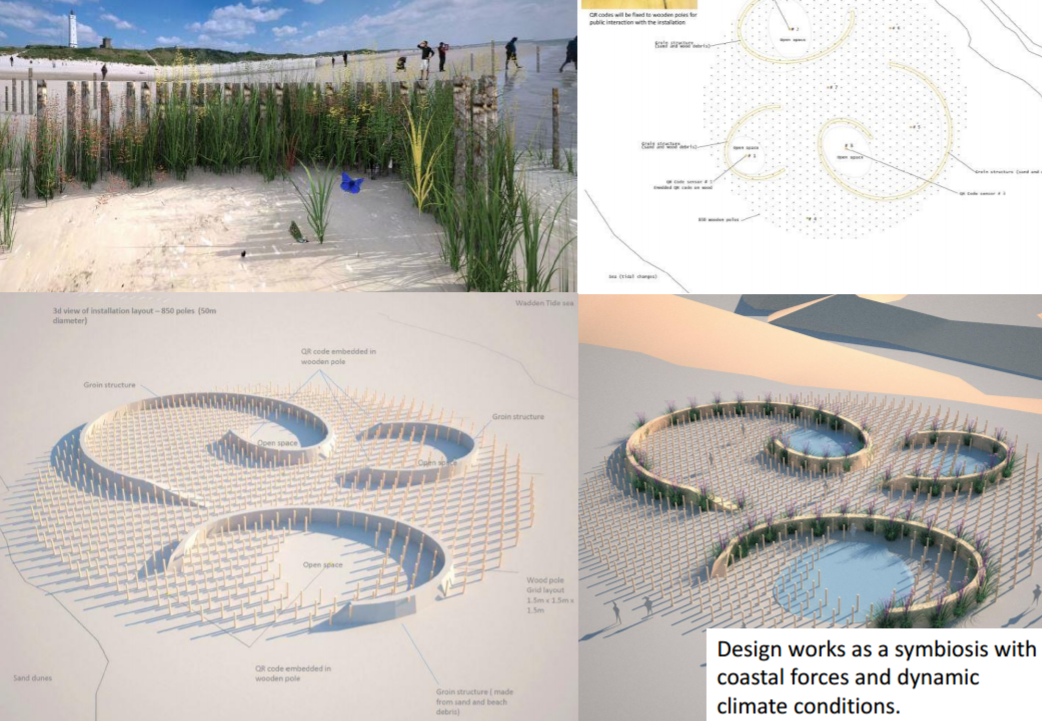
CS3608 Social Media

Seminar/Lab Sentiment Analysis

# Introduction

You have used Social Media to uncover structure in social networks or communities (NodeXL). We are now going to look at specific social media posts to determine opinions. Using the R software package you are going to analyse a set of microblog posts that were made in 2016 at a Danish art installation held on the beach. Visitors moved around the exhibit and posted comments using a mobile app that was triggered by a QR code. You can find some data on BBL that will be used for this analysis (download this onto your PC and take a look at it).

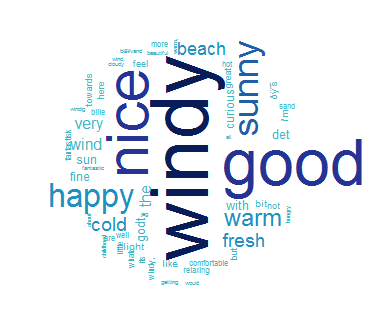


We will be creating a number of visualisations today, you should reflect later on using lab session techniques as a means to analyse customer opinions. For example, could you use social media to calculate emotion at stages of a customer journey (thinking of buying, buying, post purchase).

The lab aims to cover a number of activities: 1) Introductory use of R and the Syuzhet Package (see <ftp://cran.r-project.org/pub/R/web/packages/syuzhet/vignettes/syuzhet-vignette.html>) 2) basic data analysis and 3) emotion detection. Throughout the lab, try to think about how data and specific visualisations can be used to better understand a market, customer or product experience.

# A quick and easy wordcloud

Word clouds are a useful 1st visualisation in many cases. Load the wadden.csv file (**copy file from Blackboard to h:\wadden.csv**) and review the text microblogs in an editor or Excel. Create a wordcloud using one of the online tools (e.g. https://tagul.com/)



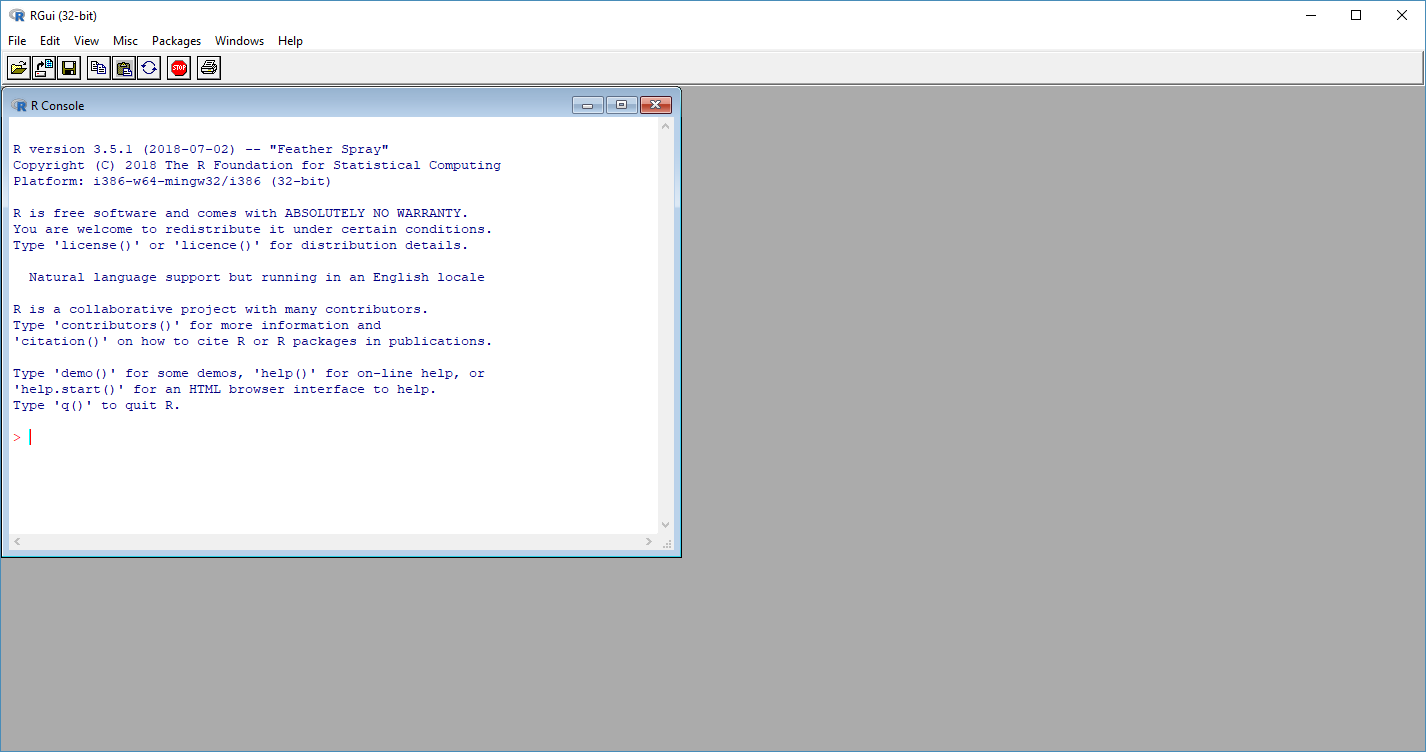
How many objects and variables were loaded (see Output – Environment) - explore some of the data. This WordCloud gives you some basic insight into what people are posting, using the ever popular word frequency. Why not change the colours?

# R – a stats and graph package

R is a language and environment for statistical computing and graphics. Similar to many development environments, it allows the user to add specific packages (which in our case will be sentiment analysis). **The same analysis carried out today can be used with a CSV file of tweets extracted using NodeXL.**

# Getting started

Find R 3.5.1 and start it to get the user interface:



We will be editing R code and running it in this console. But first need to make sure that all the libraries are available to use:

Try “library(ggplot2)” in the Source code window and Run.

If the package is not available, go to Packages 🡪 Install Packages and install **ggplot2**.

# Temporal analysis

The art installation was on the beach for 6 weeks and it would be interesting to see the number of posts on each day.

We will now use some new libraries:

library(ggplot2) // A data visualisation library

library(lubridate)// Helper library for dealing with dates and times

library(scales) // A scaling library for visualisation

You can install these packages (and others we will use later) by entering the following into the console window:

install.packages("ggplot2") – If not already installed

install.packages("lubridate")

install.packages("scales")

install.packages("syuzhet")

install.packages("reshape2")

install.packages("dplyr")

Run the code below to generate a chart of post numbers.

library(ggplot2)

library(lubridate)

library(scales)

tweets <- read.csv("C:/wadden.csv", stringsAsFactors = FALSE)

tweets$date <- dmy(tweets$date)

tweets$time <- hms(tweets$date)

ggplot(data = tweets, aes(x = date)) +

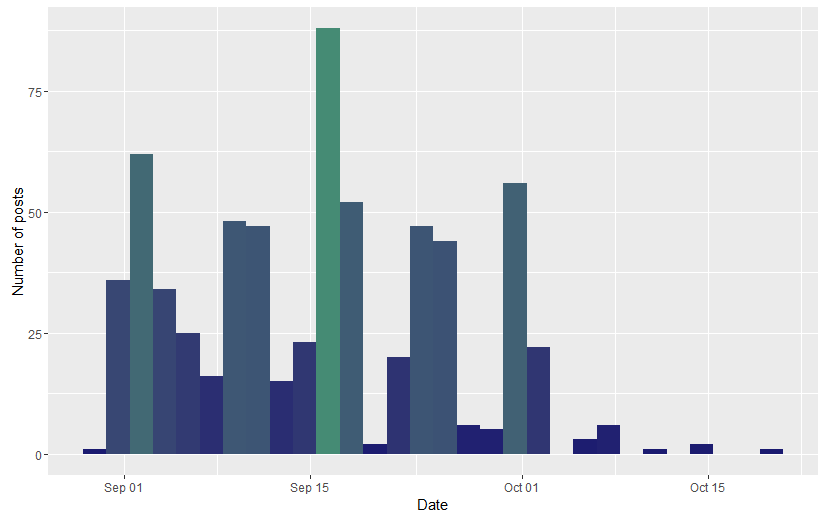
geom\_histogram(aes(fill = ..count..)) +

theme(legend.position = "none") +

xlab("Date") + ylab("Number of posts") +

scale\_fill\_gradient(low = "midnightblue", high = "aquamarine4")

As you can see, it is pretty simple to produce this plot:



Why do numbers vary so much?

# Spatial analysis

QR codes were placed at different points on the beach.



It would be nice to analyse where people are interacting with the installation. If you look at the Tweets data structure (Environment 🡪 Tweets at the top of the Output window) – what variables will be useful?

Try this plot now:

ggplot(data = tweets, aes(x = location)) +

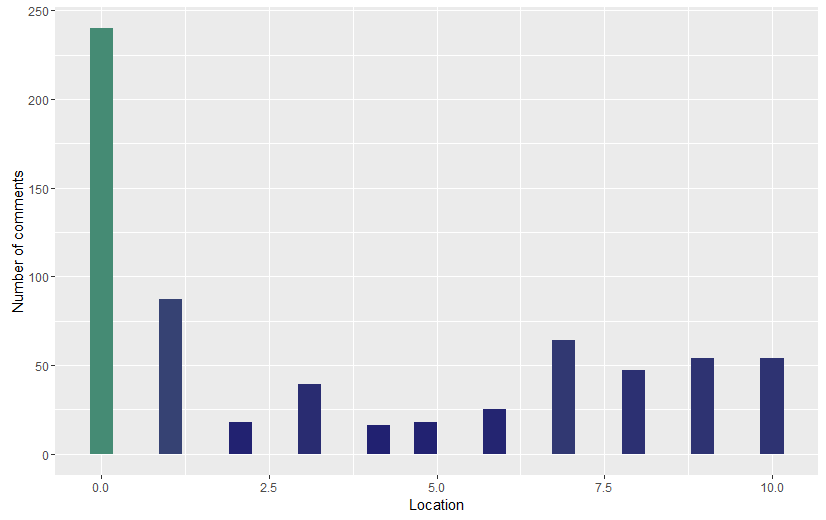
geom\_histogram(aes(fill = ..count..)) +

theme(legend.position = "none") +

xlab("Location") + ylab("Number of comments") +

scale\_fill\_gradient(low = "midnightblue", high = "aquamarine4")

The output should look like this:



# Sentiment and Emotion

We have some interesting basic analysis about how visitors interacted with the site. But it would be nice to get their opinion! The Syuznhet library has a number sentiment methods – get\_sentiment and get\_nrc\_sentiment. Type ?*method\_name* into the console window.

Let’s first start with some simple analysis of text, using the Syuznhet library – get\_nrc\_method(). Try this:

library(syuzhet)

get\_nrc\_sentiment("I am happy to get a great grade")

get\_nrc\_sentiment("It is scary when I lose my work")

How many emotions are detected in each sentence? You should be able to see this in the Console window.

Now we can look at the sentiment in our file of posts. Run the following code:

library(syuzhet)

library(lubridate)

library(ggplot2)

library(scales)

library(reshape2)

library(dplyr )

tweets <- read.csv("C:/wadden.csv", stringsAsFactors = FALSE)

**mySentiment <- get\_nrc\_sentiment(tweets$text)**

**tweets <- cbind(tweets, mySentiment)**

sentimentTotals <- data.frame(colSums(tweets[,c(5:12)]))

names(sentimentTotals) <- "count"

sentimentTotals <- cbind("sentiment" = rownames(sentimentTotals), sentimentTotals)

rownames(sentimentTotals) <- NULL

ggplot(data = sentimentTotals, aes(x = sentiment, y = count)) +

geom\_bar(aes(fill = sentiment), stat = "identity") +

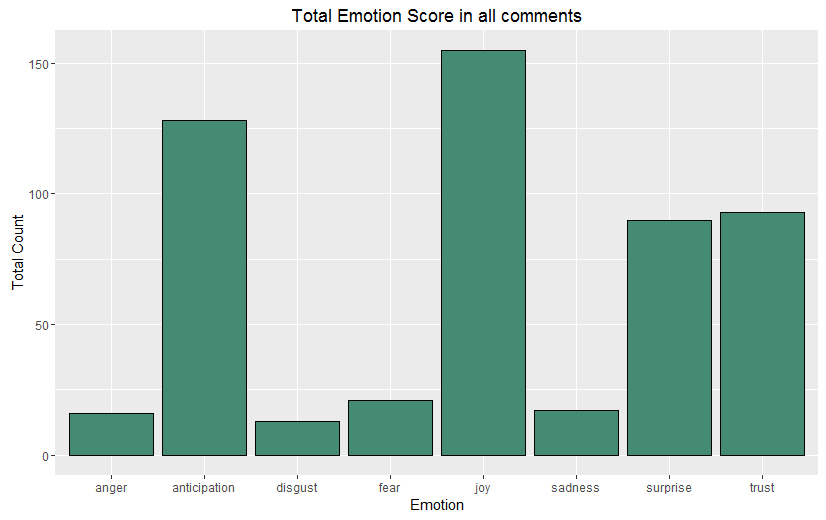
theme(legend.position = "none") +

xlab("Emotion") + ylab("Total Count") + ggtitle("Total Emotion Score in all comments") +

geom\_bar(stat="identity", fill="#458B74", colour="black")

You can see where the sentiment is calculated and how the sentiment is added to the tweets structure with cbind. Look at the environment to see if this happened.

You should get the following output:



Well done! You finished the lab.

Think about how you could use sentiment analysis for:

* Brand Health
* Manage a crisis with a product or service
* Competitive Research
* Evaluate Initiatives
* **User or experience Research (Coursework – personas & jounaeys)**

You can also find more interesting discussion on sentiment (including the library you used):

* <http://blog.revolutionanalytics.com/2016/01/twitter-sentiment.html>
* <http://www.matthewjockers.net/2016/08/11/more-syuzhet-validation/>

# INSTALLING R AND R STUDIO

If you want to install the software at home (community editions are free):

Download R from <https://cran.r-project.org/bin/windows/base/>

Download R Studio from <https://www.rstudio.com/products/rstudio/>