BIRMINGHAM CITY
University

# CMP5352 Report - TITLE NEEDED

## DRAFT VERSION

Lewis Higgins - Student ID 22133848

Word count: XXXX

**Abstract**

text
    text
    text

# Contents

# Introduction

Text text text

Text text text

aaa

# Motivation and objectives

```r
library(ggplot2)

# Side-note that if you run this as a file and not in the IDE the plots will actually be put
# into a PDF for you in the active working directory (here).

# We can use a randomly selected sample of the dataset for the graph.
# However, because the results should be reproducible, we should set an RNG seed.
set.seed(1000)

# Select the numbered rows of the numbers produced by sample. Sample picks 100 random numbers.
dsmall <- diamonds[sample(nrow(diamonds), 100),]
# The random comma at the end tells R that you're slecting ROWS, not columns.
# If you don't put this comma it assumes you're looking to select the columns.

qplot(log(carat), log(price), data = dsmall, geom = "smooth")
# Can supply multiple geoms in a vector
qplot(log(carat), log(price), data = dsmall, geom = c("point", "smooth")) # Different smooth line?

# Span varies the smoothness of geom_smooth from 0 to 1 where 1 is the smoothest.
# It states that span is an unknown parameter, yet this does actually modify the produced graph.
# 0.2 is the minimum before R throws warnings. 0.1 works with warnings, but anything lower produces
# no smooth line. Though, using 0.1 means you might as well not even put a smooth line.
qplot(log(carat), log(price), data = dsmall, geom = c("point", "smooth"), span = 0.2)

# You can also fit a linear model to the graph via lm.
qplot(log(carat), log(price), data = dsmall, geom = c("point", "smooth"), method = "lm")

# Scatterplotting a different dataset, ggplot's builtin mpg (car fuel economy data)
qplot(displ, hwy, data = mpg, color = drv)
# If you provided a color argument to this, it would draw one smooth for every color.
qplot(displ, hwy, data = mpg, geom = c("point", "smooth"))
# Answers the question "How are engine size and fuel economy related?"
# Turning cylinder into a factor (categorical data). Basically counts the appearances of each value.
```

```r
qplot(displ, hwy, data = mpg, color = factor(cyl))
# We can use all arguments previously shown at once.
# Note that there aren't enough 5 cylinders to fit a line, so there isn't one.
qplot(displ, hwy, data = mpg, color = factor(cyl), geom = c("point", "smooth"), method = "lm")

### --- Faceting --- ###


# . acts as a placeholder, indicating that there's no variable.
# Results in three seperate histograms, one of each drive class (four-wheel, front, rear-wheel).
qplot(hwy, data = mpg, facets = drv ~ ., binwidth = 2) # Could add colors. Doesn't help much though.

# Flips sideways. displ is displacement. Air movement per engine rev possibly
qplot(displ, hwy, data = mpg, facets = . ~ drv, geom = c("point", "smooth"))

# Reusing the diamond set.
qplot(carat, data = diamonds, facets = color ~ ., geom = "histogram")

# ..density.. tells ggplot to map the density as the Y-axis, instead of just counting.
qplot(carat, ..density.., data = diamonds, facets = color ~ ., geom = "histogram")

# Plots thirty five histograms by also grouping by cut.
qplot(carat, data = diamonds, facets = color ~ cut, geom = "histogram")
```

aaaa

# Experimental results

text text text
    text text text

# Summary & conclusion

text text text

text text text