End of Course Summative Survey Analysis

Author: Cheryl Calhoun

Last Edit: 24 August, 2019

# Overview

End of course survey were conducted in all five cohorts. The survey was designed by program organizers and learners were required to complete the survey prior to entering the investor pitch competition. Theoretically, all learners completing the course should have completed the survey, however, the number of surveys available are smaller than the number of participants who are designated as completing the course.

The end of course surveys for Cohorts 2013 - 2015 were provided in the form of scanned copies of the paper surveys. Qualtrics surveys were developed that matched the items in the paper surveys and all surveys were entered. The paper surveys were numbered 1-n and this number is recorded as one of the variables in the data set. This will allow validation of data against the paper survey if needed. The end of course survey for Cohort 2016 was conducted using Survey Monkey and results were provided in electronic format.

The data from all four cohorts was merged into a single data set so that analysis can be run within each cohort and across cohorts. See “Tidying the Data”, for more information on how this data set was merged and prepared for analysis.

This analysis is based on “Descriptive Statistics for Likert Data”, *Summary and Analysis of Extension Program Evaluation in R* by Salvatore S. Mangiafico. Retrieved from <Http://rcompanion.org/handbook/E_02.html> on March 10, 2017.

# Load data file that was created in TidyData.Rmd  
load("data/Summative.Rda")  
  
# Extract descriptive data for analysis.  
Descriptive <- select(Summative, ID, Cohort, Team, Degree, PaperID,   
 Discipline, Age, Relationship, Race, Income, Employment)  
  
# Extract qualitative data for analysis.  
Qualitative <- select(Summative, ID, Cohort, Team, Degree, PaperID,   
 starts\_with("TEXT"))  
  
# Extract ordinal & likert scale data for analysis.  
Ordinal <- select(Summative, ID, Cohort, Team, Degree, Race, Age, Relationship,  
 Discipline, Employment, Income, PaperID, starts\_with("Q"))  
Ordinal$Cohort <- factor(Ordinal$Cohort, ordered=TRUE)  
  
# Extract competency (Before & After) data for analysis.  
# Extracting data for analysis  
Competency <- select(Summative, ID, Cohort, Team, Degree, Race, Age, Relationship,  
 Discipline, Employment, Income, Q36.1\_1:Q16.2\_5)  
  
# Extract 2016 data for analysis. These are questions that don't exists in prior cohorts.  
Summative2016 <- select(filter(Summative, Cohort==2016), ID, Cohort, Team,   
 Degree, PaperID, starts\_with("Q5"))  
  
# Extract teams data for analysis  
Descriptive$Team <- factor(Descriptive$Team, ordered=TRUE)  
Teams <- as.data.frame(summary(Descriptive$Team))  
colnames(Teams) <- c("Total")  
rownames(Teams) <- c(levels(Descriptive$Team), "NA")  
  
# Establish Ordinal Lists  
ScaleL = c("1", "2", "3", "4", "5")  
DegreeL = c("HS", "Associate", "Bachelor", "Master","PhD", "Other")  
CohortL <- c("2013", "2014", "2015", "2016")  
TCohorts <- c("Totals", "2013", "2014", "2015", "2016")  
TimeL <- c("Too Short", "About Right", "Too Long")

# Demographic Analysis

This section will provide a demographic analysis of Cohorts (years), including team (mentor) assignment and maximum earned degree. In 2015, questions were added to collect information about Discipline, Age, Relationship, Ethnicity, Income, and Employment.

* There are a total of 151 respondents.
* There are a total of 31 teams.
* The average size of each team is 4.87.
* The largest team is 7.
* The smallest team is 0.
* The number of respondants per Cohort is 33, 34, 42, 42.

## Education

The program specifically recruited women with advanced degrees. While there are some women who participated with only a HS diploma, those women were specifically selected because of their prior knowledge of and experience in entrepreneurship. Degree will be used as an independent variable in question analysis to determine if there is any correlation between degree level and program experience as reported in the end of course survey. *Note: This question is also included on the learner application.*

**What is the highest level of education completed? (Circle One)**

# Extract the Degree information from the Learners data frame.  
Degrees <- select(Descriptive, Cohort, Degree)  
Degrees <- Degrees[complete.cases(Degrees),]  
Degrees$Cohort <- factor(Degrees$Cohort, ordered=TRUE)  
Degrees <- as.data.frame(Degrees)  
# Find the total number of degrees for the applicants (note NA's should be the 2012 applicants).  
Totals = summary(Degrees$Degree)  
nCohort = summary(Degrees$Cohort)  
  
# Find the total number of degrees by cohort.  
XT = table(Degrees$Cohort, Degrees$Degree, useNA="ifany")  
XT = rbind(XT, Totals)  
XT = prop.table(XT, margin = 1)  
XT <- apply(XT\*100, 2, function(u) sprintf("%.0f%%", u))  
row.names(XT) <- c("2013", "2014", "2015", "2016", "Totals")  
  
kable(XT, padding = 4, caption=" ")

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | HS | Associate | Bachelor | Master | PhD | Other |
| 2013 | 3% | 7% | 27% | 33% | 27% | 3% |
| 2014 | 0% | 11% | 7% | 50% | 32% | 0% |
| 2015 | 0% | 0% | 28% | 44% | 28% | 0% |
| 2016 | 7% | 0% | 45% | 29% | 19% | 0% |
| Totals | 3% | 4% | 29% | 38% | 26% | 1% |

## Disclipline/Expertise

The program attempted to recruit women from a variety of disciplines. This was so that program teams would have a variety of expertise that would facilitate the entire business development process. The data shows the highest percentage of learners come from science and communications disciplines.

This question was first asked in 2015. There are some non-respondents and therefore some NAs in this response. This data will be used as an independent variable in question analysis to determine if there is any correlation between discipline level and program experience as reported in the end of course survey.

**What is your area of expertise? (Circle One)**

# Create Disciplines dataframe  
Disciplines <- select(Descriptive, Cohort, Discipline)  
Disciplines <- Disciplines[complete.cases(Disciplines),]  
Disciplines <- filter(Disciplines, Discipline != "")  
Disciplines$Cohort <- factor(Disciplines$Cohort, ordered=TRUE)  
Disciplines$Discipline <- factor(Disciplines$Discipline, ordered=TRUE)  
Disciplines <- as.data.frame(Disciplines)  
  
# Create Disciplines cross table with Cohort  
df <- bind\_cols(  
 as.data.frame(summary(na.omit(Disciplines$Discipline))),  
 as.data.frame.matrix(table(Disciplines$Discipline, Disciplines$Cohort, useNA = "no"))  
 )  
  
# Create the proportions table  
proportions <- matrix(NA, nrow=nrow(df), ncol=ncol(df))  
proportions <- as.data.frame(proportions)  
  
for(i in 1:nrow(df)) {  
 for(j in 1:ncol(df)) {  
 proportions[i,j] <- df[i,j] / colSums(df[j])  
 }  
}  
  
# Ensure data table is a data frame, add column names and row names, and format for percentages.  
colnames(df) <- c("Total", 2015:2016)  
rownames(df) <- c(levels(Disciplines$Discipline))  
Results <- as.data.frame(apply(proportions\*100, 2, function(u) sprintf("%.0f%%", u)))  
colnames(Results) <- c("Total", 2015:2016)  
rownames(Results) <- c(levels(Disciplines$Discipline))  
  
# Display the results.  
kable(Results, padding = 4, caption=" ")

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total | 2015 | 2016 |
| Business | 13% | 12% | 14% |
| Computer/IT | 9% | 8% | 11% |
| Engineering | 17% | 10% | 24% |
| Finance | 5% | 8% | 3% |
| Marketing/Communications/Design | 19% | 25% | 14% |
| Other | 3% | 5% | 0% |
| Science | 34% | 32% | 35% |

## Age

The program attempted to recruit women with a variety of experience and thus it could be speculated that the ages of the learners would be distributed. Because of the way this question was asked, we are restricted to analyzing the ordinal responses. If this data had been collected as a numerical response we would have been able to calculate means and other descriptive statistics on the age data.

**What is your age? (Circle one)**

# Create Age dataframe  
Age <- select(Descriptive, Cohort, Age)  
Age <- Age[complete.cases(Age),]  
Age <- filter(Age, Age != "")  
Age$Cohort <- factor(Age$Cohort, ordered=TRUE)  
Age$Age <- factor(Age$Age, ordered=TRUE)  
Age <- as.data.frame(Age)  
  
# Create Age cross table with Cohort  
df <- bind\_cols(  
 as.data.frame(summary(na.omit(Age$Age))),  
 as.data.frame.matrix(table(Age$Age, Age$Cohort, useNA = "no"))  
 )  
  
# Create the proportions table  
proportions <- matrix(NA, nrow=nrow(df), ncol=ncol(df))  
proportions <- as.data.frame(proportions)  
  
for(i in 1:nrow(df)) {  
 for(j in 1:ncol(df)) {  
 proportions[i,j] <- df[i,j] / colSums(df[j])  
 }  
}  
  
# Ensure data table is a data frame, add column names and row names, and format for percentages.  
Results <- as.data.frame(apply(proportions\*100, 2, function(u) sprintf("%.0f%%", u)))  
colnames(Results) <- c("Total", "2015", "2016")  
rownames(Results) <- c(levels(Age$Age))  
  
# Display the results.  
kable(Results, padding = 4, caption=" ")

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total | 2015 | 2016 |
| 18-24 | 16% | 15% | 17% |
| 25-34 | 34% | 32% | 36% |
| 35-44 | 23% | 22% | 24% |
| 45-54 | 18% | 20% | 17% |
| 55-64 | 9% | 10% | 7% |

## Family Status

**Circle the answer that best describes your current situation.**

# Create the Status dataframe  
Status <- select(Descriptive, Cohort, Relationship)  
Status <- Status[complete.cases(Status),]  
Status <- filter(Status, Relationship != "")  
Status$Cohort <- factor(Status$Cohort, ordered=TRUE)  
Status$Relationship <- factor(Status$Relationship, ordered=TRUE)  
Status <- as.data.frame(Status)  
  
# Create Age cross table with Cohort  
df <- bind\_cols(  
 as.data.frame(summary(na.omit(Status$Relationship))),  
 as.data.frame.matrix(table(Status$Relationship, Status$Cohort, useNA = "no"))  
 )  
  
# Create the proportions table  
proportions <- matrix(NA, nrow=nrow(df), ncol=ncol(df))  
proportions <- as.data.frame(proportions)  
  
for(i in 1:nrow(df)) {  
 for(j in 1:ncol(df)) {  
 proportions[i,j] <- df[i,j] / colSums(df[j])  
 }  
}  
  
# Ensure data table is a data frame, add column names and row names, and format for percentages.  
Results <- as.data.frame(apply(proportions\*100, 2, function(u) sprintf("%.0f%%", u)))  
colnames(Results) <- c("Total", "2015", "2016")  
rownames(Results) <- c(levels(Status$Relationship))  
  
# Display the results.  
kable(Results, padding = 4, caption=" ")

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total | 2015 | 2016 |
| married or in a committed relationship with grown children (18+) | 15% | 18% | 12% |
| married or in a committed relationship with no children | 29% | 21% | 36% |
| married or in a committed relationship with school aged children (5-18) | 12% | 13% | 12% |
| married or in a committed relationship with younger child/ren (under 5) | 5% | 8% | 2% |
| single parent with grown children (18+) | 1% | 0% | 2% |
| single with no children | 38% | 39% | 36% |

## Ethnicity/Race

**Which of the following best represents your racial or ethnic heritage?**

# Create the Ethnicity dataframe  
Ethnicity <- select(Descriptive, Cohort, Race)  
Ethnicity <- Ethnicity[complete.cases(Ethnicity),]  
Ethnicity <- filter(Ethnicity, Race != "")  
Ethnicity$Cohort <- factor(Ethnicity$Cohort, ordered=TRUE)  
Ethnicity$Race <- factor(Ethnicity$Race, ordered=TRUE)  
Ethnicity <- as.data.frame(Ethnicity)  
  
# Create Age cross table with Cohort  
df <- bind\_cols(  
 as.data.frame(summary(na.omit(Ethnicity$Race))),  
 as.data.frame.matrix(table(Ethnicity$Race, Ethnicity$Cohort, useNA = "no"))  
 )  
  
# Create the proportions table  
proportions <- matrix(NA, nrow=nrow(df), ncol=ncol(df))  
proportions <- as.data.frame(proportions)  
  
for(i in 1:nrow(df)) {  
 for(j in 1:ncol(df)) {  
 proportions[i,j] <- df[i,j] / colSums(df[j])  
 }  
}  
  
# Ensure data table is a data frame, add column names and row names, and format for percentages.  
Results <- as.data.frame(apply(proportions\*100, 2, function(u) sprintf("%.0f%%", u)))  
colnames(Results) <- c("Total", "2015", "2016")  
rownames(Results) <- c(levels(Ethnicity$Race))  
  
# Display the results.  
kable(Results, padding = 4, caption=" ")

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total | 2015 | 2016 |
| Black, Afro-Caribbean, or African American | 5% | 6% | 5% |
| East Asian or Asian American | 11% | 6% | 15% |
| Latino or Hispanic American | 14% | 9% | 20% |
| Middle Eastern or Arab American | 3% | 3% | 2% |
| Non-Hispanic White or Euro-American | 61% | 69% | 54% |
| Other | 4% | 9% | 0% |
| South Asian or Indian American | 3% | 0% | 5% |

## Household Income

**What was your total household income before taxes during the past 12 months?**

# Create income dataframe  
Income <- select(Descriptive, Cohort, Income)  
Income <- Income[complete.cases(Income),]  
Income <- filter(Income, Income != "")  
Income$Cohort <- factor(Income$Cohort, ordered=TRUE)  
Income$Income <- factor(Income$Income, ordered=TRUE)  
Income <- as.data.frame(Income)  
  
# Create Age cross table with Cohort  
df <- bind\_cols(  
 as.data.frame(summary(na.omit(Income$Income))),  
 as.data.frame.matrix(table(Income$Income, Income$Cohort, useNA = "no"))  
 )  
  
# Create the proportions table  
proportions <- matrix(NA, nrow=nrow(df), ncol=ncol(df))  
proportions <- as.data.frame(proportions)  
  
for(i in 1:nrow(df)) {  
 for(j in 1:ncol(df)) {  
 proportions[i,j] <- df[i,j] / colSums(df[j])  
 }  
}  
  
# Ensure data table is a data frame, add column names and row names, and format for percentages.  
Results <- as.data.frame(apply(proportions\*100, 2, function(u) sprintf("%.0f%%", u)))  
colnames(Results) <- c("Total", "2015", "2016")  
rownames(Results) <- c(levels(Income$Income))  
  
# Display the results.  
kable(Results, padding = 4, caption=" ")

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total | 2015 | 2016 |
| Less than $25,000 | 22% | 23% | 21% |
| $25,000 to $34,999 | 4% | 3% | 5% |
| $35,000 to $49,999 | 11% | 13% | 10% |
| $50,000 to $74,999 | 17% | 17% | 17% |
| $75,000 to $99,999 | 17% | 17% | 17% |
| $100,000 to $149,999 | 19% | 20% | 19% |
| $150,000 or more | 10% | 7% | 12% |

## Employment

**Please circle the option(s) that best describe(s) your current situation. OK to choose more than one if applicable.**

# Create employment dataframe  
Employment <- select(Descriptive, Cohort, Employment)  
Employment <- Employment[complete.cases(Employment),]  
Employment <- filter(Employment, Employment != "")  
Employment$Cohort <- factor(Employment$Cohort, ordered=TRUE)  
Employment$Employment <- factor(Employment$Employment, ordered=TRUE)  
Employment <- as.data.frame(Employment)  
  
# Create Age cross table with Cohort  
df <- bind\_cols(  
 as.data.frame(summary(na.omit(Employment$Employment))),  
 as.data.frame.matrix(table(Employment$Employment, Employment$Cohort, useNA = "no"))  
 )  
  
# Create the proportions table  
proportions <- matrix(NA, nrow=nrow(df), ncol=ncol(df))  
proportions <- as.data.frame(proportions)  
  
for(i in 1:nrow(df)) {  
 for(j in 1:ncol(df)) {  
 proportions[i,j] <- df[i,j] / colSums(df[j])  
 }  
}  
  
# Ensure data table is a data frame, add column names and row names, and format for percentages.  
Results <- as.data.frame(apply(proportions\*100, 2, function(u) sprintf("%.0f%%", u)))  
colnames(Results) <- c("Total", "2015", "2016")  
rownames(Results) <- c(levels(Employment$Employment))  
  
# Display the results.  
kable(Results, padding = 4, caption=" ")

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total | 2015 | 2016 |
| Employed at a non-technology non-start up company | 7% | 0% | 12% |
| Employed at a non-technology startup | 11% | 6% | 14% |
| Employed at a technology non-startup company | 12% | 16% | 10% |
| Employed at a technology startup | 7% | 6% | 7% |
| Master’s student | 11% | 9% | 12% |
| MBA student | 3% | 3% | 2% |
| Owned my own non-technology business | 3% | 0% | 5% |
| Owned my own technology business | 12% | 19% | 7% |
| PhD student | 11% | 3% | 17% |
| Postdoc | 12% | 19% | 7% |
| Unemployed (not a student) | 5% | 6% | 5% |
| Work part-time (not a student) | 1% | 0% | 2% |
| Work part-time (not a student) | 5% | 12% | 0% |

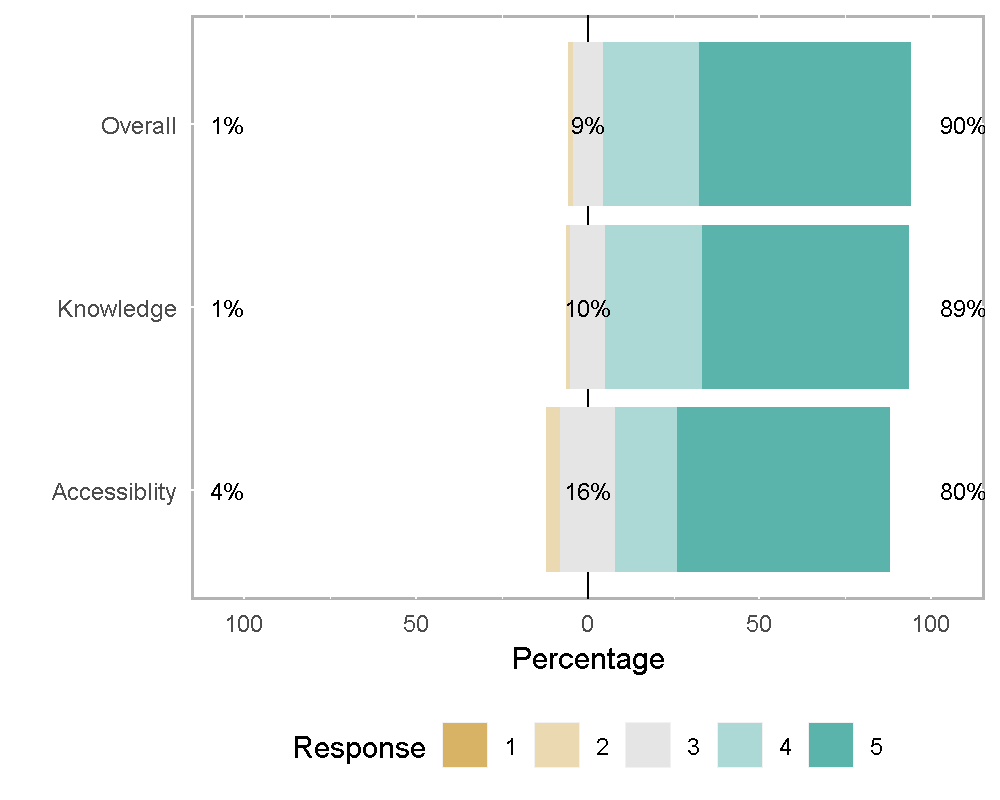
# Mentor Feedback

This section consists of three questions based on the team mentor assignment. Respondents self-identified their mentor and then answered the Likert scale question, followed by an open ended qualitative response.

* Q7: Who was your mentor?
  + Knowledge: How would you rate your mentor’s knowledge regarding entrepreneurship? (Likert-Scale (1=poor, 5=Excellent))
  + Accessiblity: How would you rate your mentor’s accessibility outside of the scheduled program meeting hours? (Likert-Scale (1=poor, 5=Excellent))
  + Overall: How would you rate your mentor overall? (Likert-Scale (1=poor, 5=Excellent)
  + TEXT8: Please provide any additional feedback regarding the mentorship experience.

**Responses Across all Cohorts**

# Extracting data for analysis  
Data <- select(Ordinal, Knowledge=Q7\_1, Accessiblity=Q7\_2, Overall=Q7\_3)  
  
## Create ordered factors of Likert data, otherwise R will alphabetize them.  
Data$Knowledge = factor(Data$Knowledge, ordered = TRUE, levels = ScaleL)  
Data$Accessiblity = factor(Data$Accessiblity, ordered = TRUE, levels = ScaleL)  
Data$Overall = factor(Data$Overall, ordered = TRUE, levels = ScaleL)  
  
## Verifying the data frame  
#headTail(Data)  
#str(Data)  
# summary(Data)  
  
Result = likert(Data)  
  
## Bar Ploat  
plot(Result,   
 type="bar",  
 main = "Mentor Feedback by Question",  
 xlab = "Question",  
 ylab = "Responses by Percentage")



# Preparing data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Knowledge=Q7\_1, Accessiblity=Q7\_2, Overall=Q7\_3)  
  
# Combining Q7\_1, Q7\_2, Q7\_2 into one column  
Data <- Data %>%  
 gather(`Knowledge`, `Accessiblity`, `Overall`, key = "Question", value = "Likert")  
  
# Create ordered factors of Likert data  
Data$Likert.f = factor(Data$Likert,   
 ordered = TRUE,   
 levels = ScaleL)  
  
Data$Team = factor(Data$Team,  
 levels=unique(Data$Team))  
  
Data$Question = factor(Data$Question,  
 levels=unique(Data$Question))  
  
Data$ID = factor(Data$ID,  
 levels=unique(Data$ID))

**Responses Summarized by Question**

# Summary treating Likert data as numeric data  
Summarize(Likert ~ Question,  
 data=Data,  
 digits = 3)

## Question n nvalid mean sd min Q1 median Q3 max  
## 1 Knowledge 151 149 4.5 0.73 2 4 5 5 5  
## 2 Accessiblity 151 150 4.4 0.89 2 4 5 5 5  
## 3 Overall 151 149 4.5 0.71 2 4 5 5 5

**Responses Summarized by Cohort**

Summarize(Likert ~ Cohort,  
 data = Data,  
 digits = 3)

## Cohort n nvalid mean sd min Q1 median Q3 max  
## 1 1 99 94 4.2 0.93 2 4 4.5 5 5  
## 2 2 102 102 4.6 0.70 3 4 5.0 5 5  
## 3 3 126 126 4.4 0.76 2 4 5.0 5 5  
## 4 4 126 126 4.6 0.72 2 4 5.0 5 5

**Responses Summarized by Team/Mentor**

Summarize(Likert ~ Team,  
 data=Data,  
 digits=3)

## Team n nvalid mean sd min Q1 median Q3 max  
## 1 2015TC 18 18 4.0 0.97 2 4.0 4.0 5.0 5  
## 2 2015TF 18 18 4.3 0.69 3 4.0 4.0 5.0 5  
## 3 2015TA 18 18 4.2 0.81 3 4.0 4.0 5.0 5  
## 4 2015TD 21 21 4.8 0.51 3 5.0 5.0 5.0 5  
## 5 2015TG 18 18 4.9 0.32 4 5.0 5.0 5.0 5  
## 6 2015TB 15 15 4.1 0.74 3 4.0 4.0 5.0 5  
## 7 2015TE 15 15 4.3 0.80 3 4.0 4.0 5.0 5  
## 8 2014TG 15 15 4.8 0.41 4 5.0 5.0 5.0 5  
## 9 2014TA 12 12 4.0 0.85 3 3.0 4.0 5.0 5  
## 10 2014TU 3 3 4.0 0.00 4 4.0 4.0 4.0 4  
## 11 2014TC 15 15 4.7 0.72 3 5.0 5.0 5.0 5  
## 12 2014TF 15 15 5.0 0.00 5 5.0 5.0 5.0 5  
## 13 2014TB 12 12 4.3 0.89 3 3.8 5.0 5.0 5  
## 14 2014TE 12 12 4.4 0.90 3 3.8 5.0 5.0 5  
## 15 2014TD 18 18 4.7 0.46 4 4.2 5.0 5.0 5  
## 16 2013TG 12 12 4.9 0.29 4 5.0 5.0 5.0 5  
## 17 2013TU 6 6 5.0 0.00 5 5.0 5.0 5.0 5  
## 18 2013TF 18 18 4.4 0.70 3 4.0 4.5 5.0 5  
## 19 2013TE 9 8 3.9 1.13 2 3.0 4.0 5.0 5  
## 20 2013TA 18 17 3.9 0.66 3 4.0 4.0 4.0 5  
## 21 2013TC 9 9 4.3 0.71 3 4.0 4.0 5.0 5  
## 22 2013TB 9 9 4.1 1.36 2 3.0 5.0 5.0 5  
## 23 2013TD 18 15 3.6 1.12 2 3.0 3.0 5.0 5  
## 24 2016TA 18 18 4.3 0.96 2 4.0 4.5 5.0 5  
## 25 2016TE 21 21 4.9 0.36 4 5.0 5.0 5.0 5  
## 26 2016TD 18 18 4.5 0.51 4 4.0 4.5 5.0 5  
## 27 2016TF 21 21 4.9 0.48 3 5.0 5.0 5.0 5  
## 28 2016TB 15 15 4.6 0.91 2 5.0 5.0 5.0 5  
## 29 2016TC 15 15 3.9 0.80 3 3.0 4.0 4.5 5  
## 30 2016TG 18 18 4.8 0.55 3 5.0 5.0 5.0 5

**Analysis of Variance**  
From the anova results we can see that Team does have a statistically significant effect on the learners responses. This could be assumed at this stage because the question is about the team mentor. However, what we don’t know if this response is affected by other independent variables. Now we will look at Cohort, education, and income to see if these variables have an affect on learners responses.

The first analysis is for all four cohorts. As we can see, both Cohort and Team assignments are statistically significant at the p < .05 level.

# Preparing data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Degree, Race, Age, Discipline, Employment, Income,  
 Relationship, Q7\_1:Q7\_3)  
Data <- mutate(Data, sum=Q7\_1+Q7\_2+Q7\_3)  
  
# Checking for statistically significant differences using Anova  
## This analysis is for all 4 cohorts  
Result <- anova(lm(sum~Cohort+Team+Degree,data=Data))  
print(Result)

## Analysis of Variance Table  
##   
## Response: sum  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Cohort 3 33 11.11 3.28 0.024 \*  
## Team 24 142 5.91 1.74 0.029 \*  
## Degree 5 19 3.80 1.12 0.354   
## Residuals 102 345 3.39   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

This analysis is for 2015-2016 using the additional demographic indicators collected beginning in 2015.

# Checking for statistically significant differences using Anova  
## This analysis is for cohors 2015 & 2016 because new demographic indicators were added in 2015  
Data <- filter(Data, Cohort>2014)  
Result <- anova(lm(sum~Cohort+Team+Degree+Age+Race+Discipline+Employment+Income+Relationship,data=Data))  
print(Result)

## Analysis of Variance Table  
##   
## Response: sum  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Cohort 1 3.1 3.06 1.02 0.35  
## Team 12 68.1 5.67 1.90 0.20  
## Degree 3 9.3 3.08 1.03 0.43  
## Age 4 3.9 0.98 0.33 0.85  
## Race 7 16.3 2.33 0.78 0.62  
## Discipline 7 36.3 5.18 1.74 0.24  
## Employment 12 40.0 3.34 1.12 0.46  
## Income 6 18.6 3.10 1.04 0.47  
## Relationship 5 9.6 1.93 0.65 0.67  
## Residuals 7 20.9 2.98

### Qualitative Responses

This question was paired with a qualitative open ended questions which asked “Please provide any additional feedback regarding the mentorship experience:” The answers were in line with the scaled responses.

## To preserve the annonymity of participants, the results from this block are hidden.   
TEXT <- select(Qualitative, Cohort, TEXT8)  
TEXT <- filter(TEXT, TEXT8>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

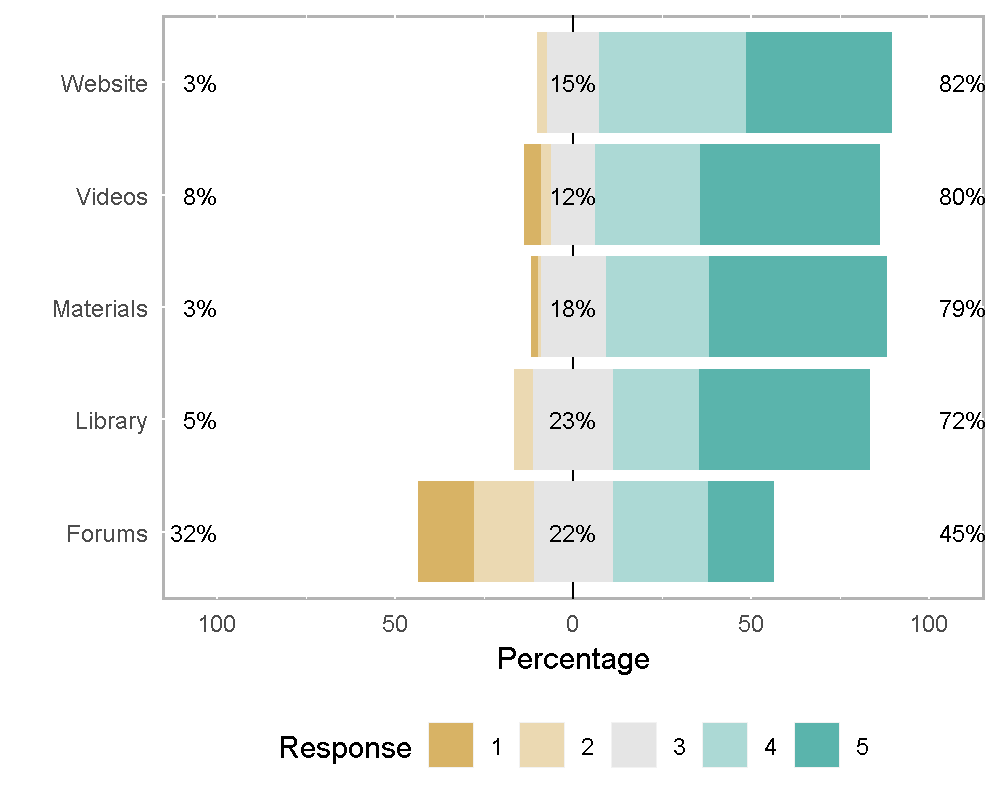
# Program Feedback

This section contains feedback about the program itself, including program resources and program content.

## Program Resources

* Website: How would you rate the program website in terms of providing information about the program?
* Library: How helpful was the UF library resources page?
* Materials: How would you rate access to the program materials on e-Learning in terms of ease of use? (Participant Workbook, Resources, Program Announcements).
* Videos: How would you rate accessing the videos on e-Learning in terms of use?
* Forums: How would you rate the Discussion Forums on e-Learning in terms of communicating with your team?

# Extracting data for analysis  
Data <- select(Ordinal, Website=Q9\_1, Library=Q9\_2, Materials=Q9\_3, Videos=Q9\_4, Forums=Q9\_5)  
  
## Create ordered factors of Likert data  
Data$Website = factor(Data$Website, ordered = TRUE, levels = ScaleL)  
Data$Library = factor(Data$Library, ordered = TRUE, levels = ScaleL)  
Data$Materials = factor(Data$Materials, ordered = TRUE, levels = ScaleL)  
Data$Videos = factor(Data$Videos, ordered = TRUE, levels = ScaleL)  
Data$Forums = factor(Data$Forums, ordered = TRUE, levels = ScaleL)  
  
# Calculate Likert Percentages  
Result = likert(Data)  
  
## Bar Ploat  
plot(Result, type="bar")



# Preparing data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Website=Q9\_1, Library=Q9\_2, Materials=Q9\_3, Videos=Q9\_4, Forums=Q9\_5)  
  
# Combining Q7\_1, Q7\_2, Q7\_2 into one column  
Data <- Data %>%  
 gather(`Website`, `Library`, `Materials`, `Videos`, `Forums`, key = "Question", value = "Likert")  
  
# Create ordered factors of Likert data  
Data$Likert.f = factor(Data$Likert, ordered = TRUE, levels = ScaleL)  
Data$Team = factor(Data$Team, levels=unique(Data$Team))  
Data$Question = factor(Data$Question, levels=unique(Data$Question))  
Data$ID = factor(Data$ID, levels=unique(Data$ID))

**Responses Summarized by Question**

# Summary treating Likert data as numeric data  
Summarize(Likert ~ Question,  
 data=Data,  
 digits = 3)

## Question n nvalid mean sd min Q1 median Q3 max  
## 1 Website 151 102 4.2 0.80 2 4 4.0 5 5  
## 2 Library 151 75 4.2 0.95 2 3 4.0 5 5  
## 3 Materials 151 148 4.2 0.92 1 4 4.5 5 5  
## 4 Videos 151 146 4.2 1.07 1 4 5.0 5 5  
## 5 Forums 151 108 3.2 1.34 1 2 3.0 4 5

**Responses Summarized by Cohort**

Summarize(Likert ~ Cohort,  
 data = Data,  
 digits = 3)

## Cohort n nvalid mean sd min Q1 median Q3 max  
## 1 1 165 140 3.8 1.17 1 3.0 4 5 5  
## 2 2 170 144 4.2 0.99 1 4.0 4 5 5  
## 3 3 210 172 4.1 1.11 1 3.8 4 5 5  
## 4 4 210 123 3.9 1.14 1 3.0 4 5 5

**Responses Summarized by Team/Mentor**

Summarize(Likert ~ Team,  
 data=Data,  
 digits=3)

## Team n nvalid mean sd min Q1 median Q3 max  
## 1 2015TC 30 26 3.5 1.53 1 3.0 4 5.0 5  
## 2 2015TF 30 20 4.2 0.59 3 4.0 4 4.2 5  
## 3 2015TA 30 28 4.0 1.05 1 3.0 4 5.0 5  
## 4 2015TD 35 26 4.4 0.94 1 4.0 5 5.0 5  
## 5 2015TG 30 25 4.1 1.15 1 3.0 5 5.0 5  
## 6 2015TB 25 23 4.0 1.11 1 3.5 4 5.0 5  
## 7 2015TE 25 19 4.2 1.03 2 4.0 5 5.0 5  
## 8 2014TG 25 21 4.3 1.02 1 4.0 5 5.0 5  
## 9 2014TA 20 14 3.7 0.99 2 3.0 4 4.0 5  
## 10 2014TU 5 5 4.0 0.00 4 4.0 4 4.0 4  
## 11 2014TC 25 23 3.8 1.11 1 3.0 4 5.0 5  
## 12 2014TF 25 19 4.5 0.61 3 4.0 5 5.0 5  
## 13 2014TB 20 19 4.3 0.75 3 4.0 4 5.0 5  
## 14 2014TE 20 17 4.5 0.87 2 4.0 5 5.0 5  
## 15 2014TD 30 26 4.1 1.28 1 3.2 5 5.0 5  
## 16 2013TG 20 19 3.9 1.49 1 3.0 5 5.0 5  
## 17 2013TU 10 10 4.7 0.48 4 4.2 5 5.0 5  
## 18 2013TF 30 27 3.6 0.97 1 3.0 4 4.0 5  
## 19 2013TE 15 12 3.0 1.13 1 2.0 3 4.0 5  
## 20 2013TA 30 22 3.8 1.15 1 3.0 4 5.0 5  
## 21 2013TC 15 15 4.1 1.33 1 4.0 5 5.0 5  
## 22 2013TB 15 12 4.2 0.94 2 4.0 4 5.0 5  
## 23 2013TD 30 23 3.8 1.11 1 3.0 4 5.0 5  
## 24 2016TA 30 18 4.4 0.85 3 4.0 5 5.0 5  
## 25 2016TE 35 21 4.2 1.08 2 3.0 5 5.0 5  
## 26 2016TD 30 16 3.7 1.30 1 3.0 4 5.0 5  
## 27 2016TF 35 21 3.2 1.48 1 2.0 4 4.0 5  
## 28 2016TB 25 15 3.7 1.05 2 3.0 4 4.5 5  
## 29 2016TC 25 14 4.4 0.51 4 4.0 4 5.0 5  
## 30 2016TG 30 18 3.9 0.87 3 3.0 4 5.0 5

**Analysis of Variance**  
From the anova results we can see that Team does have a statistically significant effect on the learners responses. This could be assumed at this stage because the question is about the team mentor. However, what we don’t know if this response is affected by other independent variables. Now we will look at Cohort, education, and income to see if these variables have an affect on learners responses.

# Preparing data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Degree, Race, Age, Discipline, Employment, Income, Relationship, Q9\_1:Q9\_5)  
Data <- mutate(Data, sum=Q9\_1+Q9\_2+Q9\_3+Q9\_4+Q9\_5)  
  
# Checking for statistically significant differences using Anova  
Result <- anova(lm(sum~Cohort+Team+Degree,data=Data))  
print(Result)

## Analysis of Variance Table  
##   
## Response: sum  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Cohort 2 0.4 0.19 0.01 0.99  
## Team 15 151.3 10.09 0.57 0.85  
## Degree 4 81.2 20.29 1.15 0.37  
## Residuals 15 263.9 17.60

# Checking for statistically significant differences using Anova  
## This analysis is for cohors 2015 & 2016 because new demographic indicators were added in 2015  
#Data <- filter(Data, Cohort>2014)  
#Result <- anova(lm(sum~Cohort+Team+Degree+Age+Race+Discipline+Employment+Income+Relationship,data=Data))  
#print(Result)

### Qualitative Responses

This question was paired with a qualitative open ended questions which asked “Please provide any additional feedback about E-Learning:” The answers were in line with the scaled responses.

## To preserve the annonymity of participants, the results from this block are hidden.   
TEXT <- select(Qualitative, Cohort, TEXT10)  
TEXT <- filter(TEXT, TEXT10>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

## Information & Technology Matching Sessions

For this analysis I will groups the Information Session & Technology Matching sessions together for analysis. While these sessions are dissemilar, the analysis is similar and the side by side comparison will allow program organizers to see the differing perceptions of the learners.

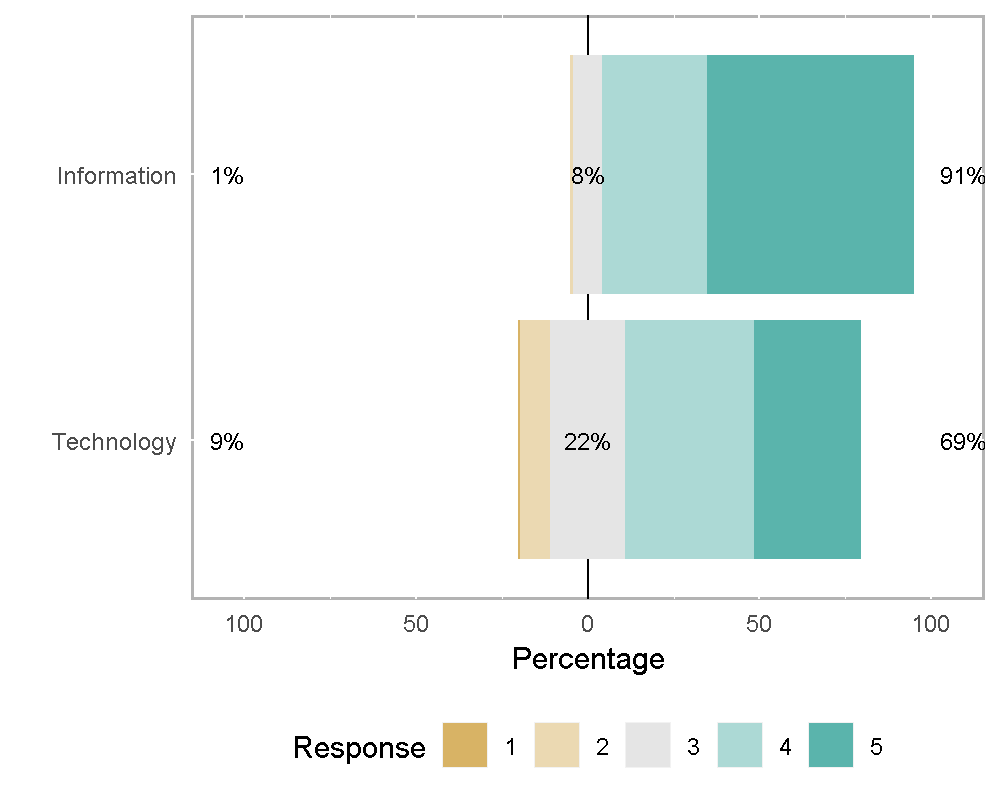
* Q1: How helpful was the Informational Session in providing an overview of what the program would encompass? (Likert-Scale (1=poor, 5=Excellent))

The Information Sessions were held in advance of learners submitting their applications. These sessions were designed to introduce potential learners to the program, give them a good idea of the content and time commitment for the program, and help them to make an informed decision about program participation.

* Q4: How would you rate your opportunity to review the technologies during the Technology Matching Session 1? (Likert-Scale (1=poor, 5=Excellent))

The first session of the educational program is a mentor matching session. During this session each mentor hosts a table. The learners are placed in random groups and participate in a “speed dating” style of mentor matching. The groups have an assigned first table and then move from table to table until they have visited all seven tables. At each table the mentor had 10 minutes to present their technology, discuss their mentoring philosophy and answer any questions. At the end of the session the learners fill out a matching request indicting their 1st, 2nd, and 3rd choices for mentor/technology team match.

# Extracting data for analysis  
Data <- select(Ordinal, Information=Q1, Technology=Q4)  
  
## Create ordered factors of Likert data  
Data$Information = factor(Data$Information, ordered = TRUE, levels = ScaleL)  
Data$Technology = factor(Data$Technology, ordered = TRUE, levels = ScaleL)  
  
# Calculate Likert Percentages  
Result = likert(Data)  
  
## Bar Ploat  
plot(Result,   
 type="bar")



# Extracting data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Information=Q1, Technology=Q4)  
  
# Combining questions into one column  
Data <- Data %>%  
 gather(`Information`, `Technology`, key = "Question", value = "Likert")  
  
## Create ordered factors of Likert data  
Data$ID = factor(Data$ID, levels=unique(Data$ID), ordered = TRUE)  
Data$Cohort = factor(Data$Cohort, levels=unique(Data$Cohort), ordered = TRUE)  
Data$Team = factor(Data$Team, levels=unique(Data$Team), ordered = TRUE)  
Data$Likert.f = factor(Data$Likert, ordered = TRUE, levels = ScaleL)  
Data$Question = factor(Data$Question, levels=unique(Data$Question))

**Responses Summarized by Question**

# Summary treating Likert data as numeric data  
Summarize(Likert ~ Question,  
 data=Data,  
 digits = 3)

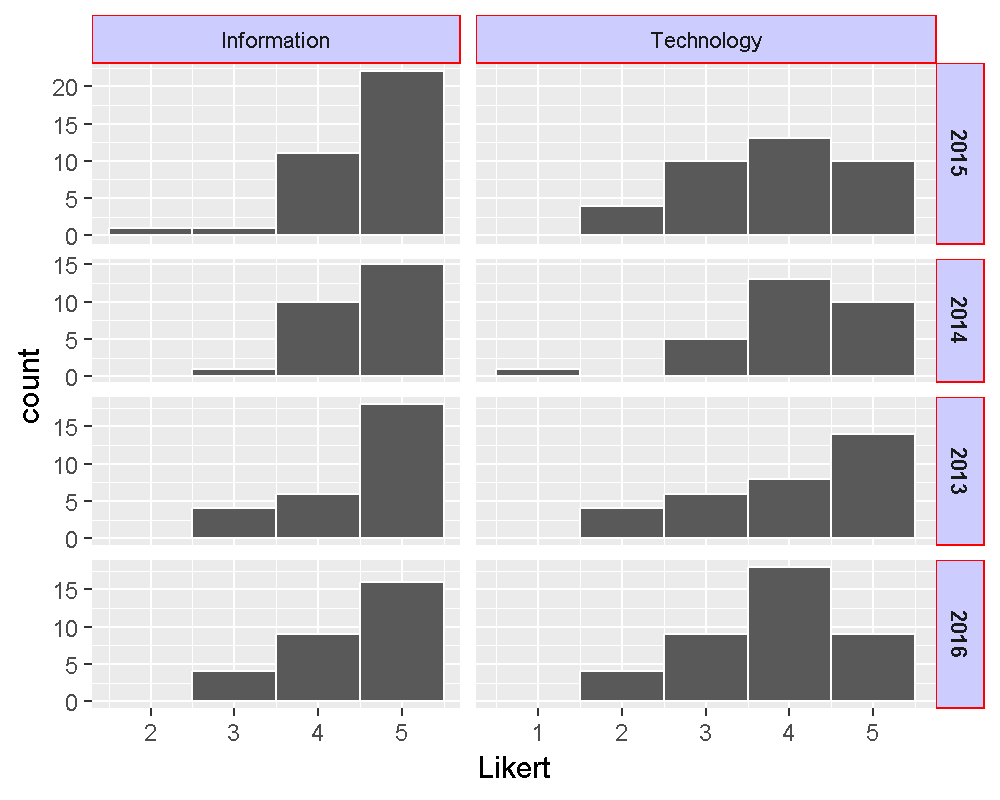
## Question n nvalid mean sd min Q1 median Q3 max  
## 1 Information 151 118 4.5 0.69 2 4 5 5 5  
## 2 Technology 151 138 3.9 0.97 1 3 4 5 5

**Responses Summarized by Cohort**

Summarize(Likert ~ Cohort + Question,  
 data=Data,  
 digits = 3)

## Cohort Question n nvalid mean sd min Q1 median Q3 max  
## 1 3 Information 42 35 4.5 0.70 2 4 5 5 5  
## 2 2 Information 34 26 4.5 0.58 3 4 5 5 5  
## 3 1 Information 33 28 4.5 0.75 3 4 5 5 5  
## 4 4 Information 42 29 4.4 0.73 3 4 5 5 5  
## 5 3 Technology 42 37 3.8 0.98 2 3 4 5 5  
## 6 2 Technology 34 29 4.1 0.92 1 4 4 5 5  
## 7 1 Technology 33 32 4.0 1.08 2 3 4 5 5  
## 8 4 Technology 42 40 3.8 0.91 2 3 4 4 5

ggplot(Data, aes(Likert)) +  
 geom\_histogram(binwidth = 1, colour = "white") +  
 facet\_grid(Cohort ~ Question, scales = "free", space="free")+  
 theme(strip.text.x = element\_text(size=8),  
 strip.text.y = element\_text(size=8, face="bold"),  
 strip.background = element\_rect(colour="red", fill="#CCCCFF"))



\*\* Analysis of Variance\*\*

# Preparing data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Degree, Race, Age, Discipline, Employment, Income,  
 Relationship, Q1:Q4)  
Data <- mutate(Data, sum=Q1+Q4)  
  
# Checking for statistically significant differences using Anova  
Result <- anova(lm(sum~Cohort+Team+Degree+Age+Race+Discipline+Employment+Income+Relationship,data=Data))

## Warning in anova.lm(lm(sum ~ Cohort + Team + Degree + Age + Race +  
## Discipline + : ANOVA F-tests on an essentially perfect fit are unreliable

print(Result)

## Analysis of Variance Table  
##   
## Response: sum  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Cohort 1 0.01 0.01   
## Team 11 14.89 1.35   
## Degree 3 1.98 0.66   
## Age 4 6.33 1.58   
## Race 7 27.80 3.97   
## Discipline 7 12.78 1.83   
## Employment 9 14.84 1.65   
## Income 2 2.57 1.29   
## Residuals 0 0.00

### Qualitative Responses Information Session

This question was paired with a qualitative open ended questions which asked “Please include feedback about the Informational Session.” The answers were in line with the scaled responses. *Note: This qualitative option was not included in the 2016 survey.*

## To preserve the annonymity of participants, the results from this block are hidden.   
TEXT <- select(Qualitative, Cohort, TEXT2)  
TEXT <- filter(TEXT, TEXT2>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

### Qualitative Responses Technology Matching Session

This question was paired with a qualitative open ended questions which asked “Please include feedback about the Technology Matching Session.” The answers were in line with the scaled responses. Some of the more descriptive responses were:

## To preserve the annonymity of participants, the results from this block are hidden.   
TEXT <- select(Qualitative, Cohort, TEXT5)  
TEXT <- filter(TEXT, TEXT5>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

## Session Video Presentations

The curriculum content for each week is delivered via pre-recorded videos. Learners are asked to rate these for Information Provided, Relevance to Project, and Speaker Presentation Skills. This three part ranking was discontinued in 2016 when learners were only asked to provide one score for each session video. To account for the variance in 3-part vs. 3-part response, this first analysis uses either the mean of the 3-part score or the single 1-part score.

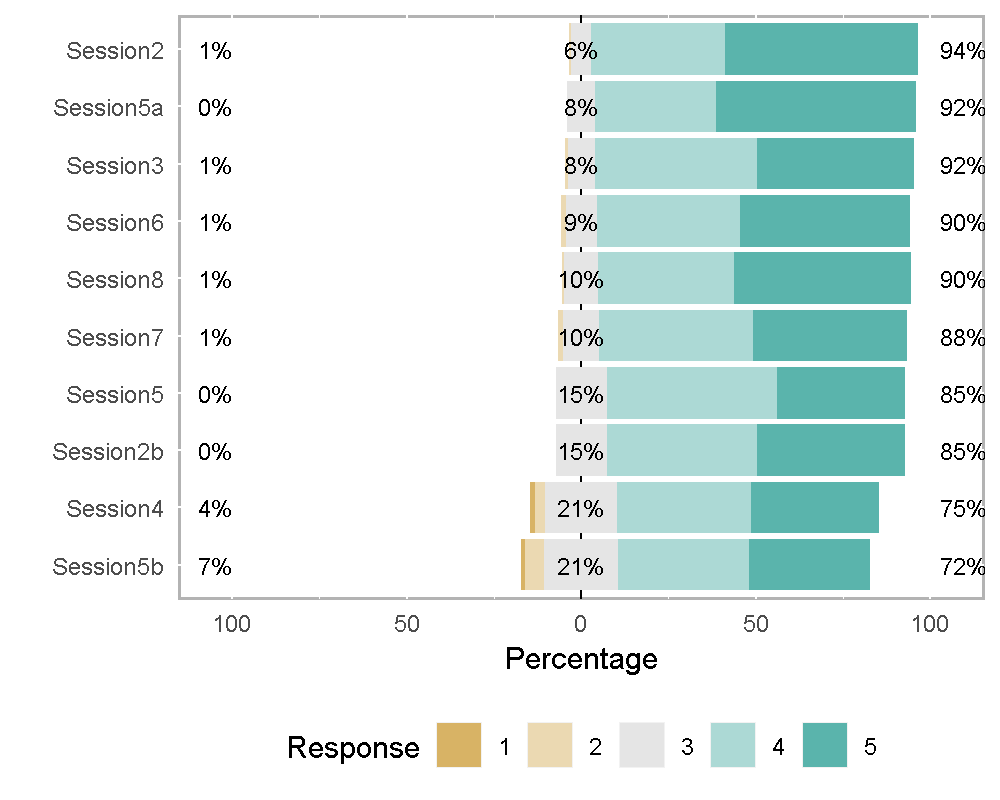
* Please rate the Videos from 1-5 for each category (scale 1=poor, 3=average, 5=excellent)
  + Q11\_1 Session 2: Understanding the Value Proposition (Randy Scott)
  + Q11\_2 Session 2b: Forming the Management Team (John Spence)
  + Q11\_3 Session 3: Market Analysis and Strategy (Karen Zaderej)
  + Q11\_4 Session 4: Commercialization Strategy (John Engels) and Intellectual Property-IP (John Byatt)
  + Q11\_5 Session 5: Financials (Angela Pate) *Was not used in 2016*
  + Q11\_6 Session 5b: Financials Spreadsheet Video (Rebecca Prince) *This video was added in 2015*
  + Q11\_7 Session 6: Company Presentations (Sue Washer)
  + Q11\_7 Session 7: Corporate Structures (John Spence)
  + Q11\_8 Session 8: Sources of Funding (Weaver Gaines)
  + Q11\_9 Session 5: Financials Spreadsheet Video (Rebecca Prince) *This video was added in 2016*

The question scores which use the three factor ranking are indicated by the “\_1“,”\_2“, and”\_3" at the end of each question. Ex.Q11\_1\_1: Information Provided, Q11\_1\_2: Relevance to Project, Q11\_1\_3: Speaker Presentation.

# Extracting data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Degree, starts\_with("Q11"))  
  
# Calculate mean score for years where Presentation, Information and relavancy were separate responses. Used raw score for 2016.  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session2 = mean(c(Q11\_1\_1, Q11\_1\_2, Q11\_1\_3, Q11\_1a), na.rm = TRUE))  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session2b = mean(c(Q11\_2\_1, Q11\_2\_2, Q11\_2\_3, Q11\_1b), na.rm = TRUE))  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session3 = mean(c(Q11\_3\_1, Q11\_3\_2, Q11\_3\_3, Q11\_2), na.rm = TRUE))  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session4 = mean(c(Q11\_4\_1, Q11\_4\_2, Q11\_4\_3, Q11\_3), na.rm = TRUE))  
  
# The video for financials switched from Angela Pate to Rebecca Prince in 2015. In 2015 both were shown, in 2016 only Rebecca Prince was shown.  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session5a = mean(c(Q11\_5\_1, Q11\_5\_2, Q11\_5\_3), na.rm = TRUE))  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session5b = mean(c(Q11\_9\_1, Q11\_9\_2, Q11\_9\_3, Q11\_4), na.rm = TRUE))  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session6 = mean(c(Q11\_6\_1, Q11\_6\_2, Q11\_6\_3, Q11\_5), na.rm = TRUE))  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session7 = mean(c(Q11\_7\_1, Q11\_7\_2, Q11\_7\_3, Q11\_6), na.rm = TRUE))  
Data <- Data %>%  
 rowwise() %>%  
 mutate(Session8 = mean(c(Q11\_8\_1, Q11\_8\_2, Q11\_8\_3, Q11\_7), na.rm = TRUE))  
  
# Select data for numeric vs ordinal evaluation  
DataS <- select(Data, ID, Cohort, Team, Degree, starts\_with("S"))  
DataL <- select(Data, starts\_with("S"))  
DataL <- as.data.frame(DataL)  
  
# Convert numberic means back to Likert values for Plot  
DataL$Session2 <- factor(as.integer(DataL$Session2), ordered=TRUE, levels=ScaleL)  
DataL$Session2b <- factor(as.integer(DataL$Session2b), ordered=TRUE, levels=ScaleL)  
DataL$Session3 <- factor(as.integer(DataL$Session3), ordered=TRUE, levels=ScaleL)  
DataL$Session4 <- factor(as.integer(DataL$Session4), ordered=TRUE, levels=ScaleL)  
DataL$Session5 <- factor(as.integer(DataL$Session5a), ordered=TRUE, levels=ScaleL)  
DataL$Session5b <- factor(as.integer(DataL$Session5b), ordered=TRUE, levels=ScaleL)  
DataL$Session6 <- factor(as.integer(DataL$Session6), ordered=TRUE, levels=ScaleL)  
DataL$Session7 <- factor(as.integer(DataL$Session7), ordered=TRUE, levels=ScaleL)  
DataL$Session8 <- factor(as.integer(DataL$Session8), ordered=TRUE, levels=ScaleL)  
  
# Calculate Likert Percentages  
Result = likert(DataL)

## Warning in likert(DataL): items parameter contains non-factors. Will  
## convert to factors

## Bar Plot  
plot(Result, type="bar", main="Session 2")



# Extracting data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Degree, starts\_with("Q11"))  
  
Information <- select(Data, ID, Cohort, Team, Degree, ends\_with("1"))  
Information <- mutate(Information, Type = "Information")  
Information <- Information %>%  
 gather(`Q11\_1\_1`, `Q11\_2\_1`, `Q11\_3\_1`, `Q11\_4\_1`, `Q11\_5\_1`, `Q11\_6\_1`, `Q11\_7\_1`,  
 `Q11\_8\_1`, `Q11\_9\_1`, key = "Question", value = "Likert" )  
  
Relevance <- select(Data, ID,Cohort, Team, Degree, ends\_with("2"))  
Relevance <- select(Relevance, -Q11\_2)  
Relevance <- mutate(Relevance, Type = "Relevance")  
Relevance <- Relevance %>%  
 gather(`Q11\_1\_2`, `Q11\_2\_2`, `Q11\_3\_2`, `Q11\_4\_2`, `Q11\_5\_2`, `Q11\_6\_2`, `Q11\_7\_2`,  
 `Q11\_8\_2`, `Q11\_9\_2`, key = "Question", value = "Likert" )  
  
Presentation <- select(Data, ID, Cohort, Team, Degree, ends\_with("3"))  
Presentation <- select(Presentation, -Q11\_3)  
Presentation <- mutate(Presentation, Type = "Presentation")  
Presentation <- Presentation %>%  
 gather(`Q11\_1\_3`, `Q11\_2\_3`, `Q11\_3\_3`, `Q11\_4\_3`, `Q11\_5\_3`, `Q11\_6\_3`, `Q11\_7\_3`,  
 `Q11\_8\_3`, `Q11\_9\_3`, key = "Question", value = "Likert" )  
  
Combined <- select(Data, ID, Cohort, Team, Degree, Q11\_1a:Q11\_7)  
Combined <- mutate(Combined, Type = "Combined")  
Combined <- Combined %>%  
 gather(`Q11\_1a`, `Q11\_1b`, `Q11\_2`, `Q11\_3`, `Q11\_4`, `Q11\_5`, `Q11\_6`, `Q11\_7`, key = "Question", value = "Likert" )  
  
Data <- bind\_rows(Information, Relevance, Presentation, Combined)

**Responses Summarized by Question**

Summarize(Likert ~ Question,  
 data=Data,  
 digits = 3)

## Question n nvalid mean sd min Q1 median Q3 max  
## 1 Q11\_1\_1 151 100 4.7 0.52 3 4.0 5.0 5 5  
## 2 Q11\_1\_2 151 100 4.7 0.49 3 4.0 5.0 5 5  
## 3 Q11\_1\_3 151 99 4.8 0.46 3 5.0 5.0 5 5  
## 4 Q11\_1a 151 42 4.2 0.81 2 4.0 4.0 5 5  
## 5 Q11\_1b 151 42 4.3 0.74 3 4.0 4.0 5 5  
## 6 Q11\_2 151 42 4.2 0.72 3 4.0 4.0 5 5  
## 7 Q11\_2\_1 151 101 4.6 0.57 3 4.0 5.0 5 5  
## 8 Q11\_2\_2 151 101 4.2 1.02 1 4.0 4.0 5 5  
## 9 Q11\_2\_3 151 100 4.7 0.55 3 4.0 5.0 5 5  
## 10 Q11\_3 151 41 3.8 1.11 1 3.0 4.0 5 5  
## 11 Q11\_3\_1 151 99 4.5 0.64 2 4.0 5.0 5 5  
## 12 Q11\_3\_2 151 99 4.7 0.55 3 4.0 5.0 5 5  
## 13 Q11\_3\_3 151 94 4.6 0.61 3 4.0 5.0 5 5  
## 14 Q11\_4 151 42 3.8 1.08 1 3.0 4.0 5 5  
## 15 Q11\_4\_1 151 93 4.4 0.72 2 4.0 5.0 5 5  
## 16 Q11\_4\_2 151 94 4.3 0.72 3 4.0 4.0 5 5  
## 17 Q11\_4\_3 151 93 4.3 0.85 1 4.0 5.0 5 5  
## 18 Q11\_5 151 40 4.0 0.88 2 3.8 4.0 5 5  
## 19 Q11\_5\_1 151 95 4.4 0.70 3 4.0 4.0 5 5  
## 20 Q11\_5\_2 151 94 4.4 0.77 2 4.0 5.0 5 5  
## 21 Q11\_5\_3 151 91 4.4 0.75 1 4.0 5.0 5 5  
## 22 Q11\_6 151 40 4.3 0.85 2 4.0 4.0 5 5  
## 23 Q11\_6\_1 151 93 4.6 0.57 3 4.0 5.0 5 5  
## 24 Q11\_6\_2 151 93 4.6 0.60 3 4.0 5.0 5 5  
## 25 Q11\_6\_3 151 92 4.7 0.46 4 4.0 5.0 5 5  
## 26 Q11\_7 151 39 4.3 0.73 3 4.0 4.0 5 5  
## 27 Q11\_7\_1 151 96 4.6 0.56 3 4.0 5.0 5 5  
## 28 Q11\_7\_2 151 95 4.2 0.81 2 4.0 4.0 5 5  
## 29 Q11\_7\_3 151 95 4.6 0.54 3 4.0 5.0 5 5  
## 30 Q11\_8\_1 151 94 4.6 0.61 2 4.0 5.0 5 5  
## 31 Q11\_8\_2 151 93 4.4 0.76 2 4.0 5.0 5 5  
## 32 Q11\_8\_3 151 92 4.6 0.63 2 4.0 5.0 5 5  
## 33 Q11\_9\_1 151 32 4.3 0.78 3 4.0 4.5 5 5  
## 34 Q11\_9\_2 151 32 4.6 0.61 3 4.0 5.0 5 5  
## 35 Q11\_9\_3 151 31 4.5 0.63 3 4.0 5.0 5 5

**Responses Sumarized by Type**

Summarize(Likert ~ Type,  
 data=Data,  
 digits = 3)

## Type n nvalid mean sd min Q1 median Q3 max  
## 1 Combined 1208 328 4.1 0.89 1 4 4 5 5  
## 2 Information 1359 803 4.5 0.63 2 4 5 5 5  
## 3 Presentation 1359 787 4.6 0.63 1 4 5 5 5  
## 4 Relevance 1359 801 4.5 0.75 1 4 5 5 5

**Responses Summarized by Question**

Summarize(Likert ~ Cohort,  
 data=Data,  
 digits = 3)

## Warning in if (tmp$Eclass != "factor") mode(lvl.lbls) <- tmp$Eclass else  
## lvl.lbls <- as.factor(lvl.lbls): the condition has length > 1 and only the  
## first element will be used

## Warning in if (storage.mode(x) == value) return(x): the condition has  
## length > 1 and only the first element will be used

## Warning in if (setSingle != isSingle) attr(x, "Csingle") <- if (setSingle)  
## TRUE: the condition has length > 1 and only the first element will be used

## Cohort n nvalid mean sd min Q1 median Q3 max  
## 1 1 1155 700 4.6 0.65 1 4 5 5 5  
## 2 2 1190 739 4.5 0.63 1 4 5 5 5  
## 3 3 1470 952 4.5 0.72 1 4 5 5 5  
## 4 4 1470 328 4.1 0.89 1 4 4 5 5

**Responses Summarized by Team**

Summarize(Likert ~ Team,  
 data=Data,  
 digits = 3,  
 na.rm = TRUE)

**Responses Summarized by Degree**

Data$Degree <- as.character(Data$Degree)  
Summarize(Likert ~ Degree,  
 data=Data,  
 digits = 3)

## Degree n nvalid mean sd min Q1 median Q3 max  
## 1 Associate 175 87 4.6 0.58 3 4 5.0 5 5  
## 2 Bachelor 1400 630 4.4 0.80 1 4 5.0 5 5  
## 3 HS 140 48 4.3 0.76 2 4 4.5 5 5  
## 4 Master 1855 1006 4.5 0.71 2 4 5.0 5 5  
## 5 Other 35 24 4.9 0.45 3 5 5.0 5 5  
## 6 PhD 1260 654 4.5 0.72 1 4 5.0 5 5

**Responses Summarized by Question & Type**

Summarize(Likert ~ Question + Type,  
 data=Data,  
 digits = 3)

## Question Type n nvalid mean sd min Q1 median Q3 max  
## 1 Q11\_1a Combined 151 42 4.2 0.81 2 4.0 4.0 5 5  
## 2 Q11\_1b Combined 151 42 4.3 0.74 3 4.0 4.0 5 5  
## 3 Q11\_2 Combined 151 42 4.2 0.72 3 4.0 4.0 5 5  
## 4 Q11\_3 Combined 151 41 3.8 1.11 1 3.0 4.0 5 5  
## 5 Q11\_4 Combined 151 42 3.8 1.08 1 3.0 4.0 5 5  
## 6 Q11\_5 Combined 151 40 4.0 0.88 2 3.8 4.0 5 5  
## 7 Q11\_6 Combined 151 40 4.3 0.85 2 4.0 4.0 5 5  
## 8 Q11\_7 Combined 151 39 4.3 0.73 3 4.0 4.0 5 5  
## 9 Q11\_1\_1 Information 151 100 4.7 0.52 3 4.0 5.0 5 5  
## 10 Q11\_2\_1 Information 151 101 4.6 0.57 3 4.0 5.0 5 5  
## 11 Q11\_3\_1 Information 151 99 4.5 0.64 2 4.0 5.0 5 5  
## 12 Q11\_4\_1 Information 151 93 4.4 0.72 2 4.0 5.0 5 5  
## 13 Q11\_5\_1 Information 151 95 4.4 0.70 3 4.0 4.0 5 5  
## 14 Q11\_6\_1 Information 151 93 4.6 0.57 3 4.0 5.0 5 5  
## 15 Q11\_7\_1 Information 151 96 4.6 0.56 3 4.0 5.0 5 5  
## 16 Q11\_8\_1 Information 151 94 4.6 0.61 2 4.0 5.0 5 5  
## 17 Q11\_9\_1 Information 151 32 4.3 0.78 3 4.0 4.5 5 5  
## 18 Q11\_1\_3 Presentation 151 99 4.8 0.46 3 5.0 5.0 5 5  
## 19 Q11\_2\_3 Presentation 151 100 4.7 0.55 3 4.0 5.0 5 5  
## 20 Q11\_3\_3 Presentation 151 94 4.6 0.61 3 4.0 5.0 5 5  
## 21 Q11\_4\_3 Presentation 151 93 4.3 0.85 1 4.0 5.0 5 5  
## 22 Q11\_5\_3 Presentation 151 91 4.4 0.75 1 4.0 5.0 5 5  
## 23 Q11\_6\_3 Presentation 151 92 4.7 0.46 4 4.0 5.0 5 5  
## 24 Q11\_7\_3 Presentation 151 95 4.6 0.54 3 4.0 5.0 5 5  
## 25 Q11\_8\_3 Presentation 151 92 4.6 0.63 2 4.0 5.0 5 5  
## 26 Q11\_9\_3 Presentation 151 31 4.5 0.63 3 4.0 5.0 5 5  
## 27 Q11\_1\_2 Relevance 151 100 4.7 0.49 3 4.0 5.0 5 5  
## 28 Q11\_2\_2 Relevance 151 101 4.2 1.02 1 4.0 4.0 5 5  
## 29 Q11\_3\_2 Relevance 151 99 4.7 0.55 3 4.0 5.0 5 5  
## 30 Q11\_4\_2 Relevance 151 94 4.3 0.72 3 4.0 4.0 5 5  
## 31 Q11\_5\_2 Relevance 151 94 4.4 0.77 2 4.0 5.0 5 5  
## 32 Q11\_6\_2 Relevance 151 93 4.6 0.60 3 4.0 5.0 5 5  
## 33 Q11\_7\_2 Relevance 151 95 4.2 0.81 2 4.0 4.0 5 5  
## 34 Q11\_8\_2 Relevance 151 93 4.4 0.76 2 4.0 5.0 5 5  
## 35 Q11\_9\_2 Relevance 151 32 4.6 0.61 3 4.0 5.0 5 5

### Additional Feedback Videos

* Please provide any additional feedback regarding the session videos:

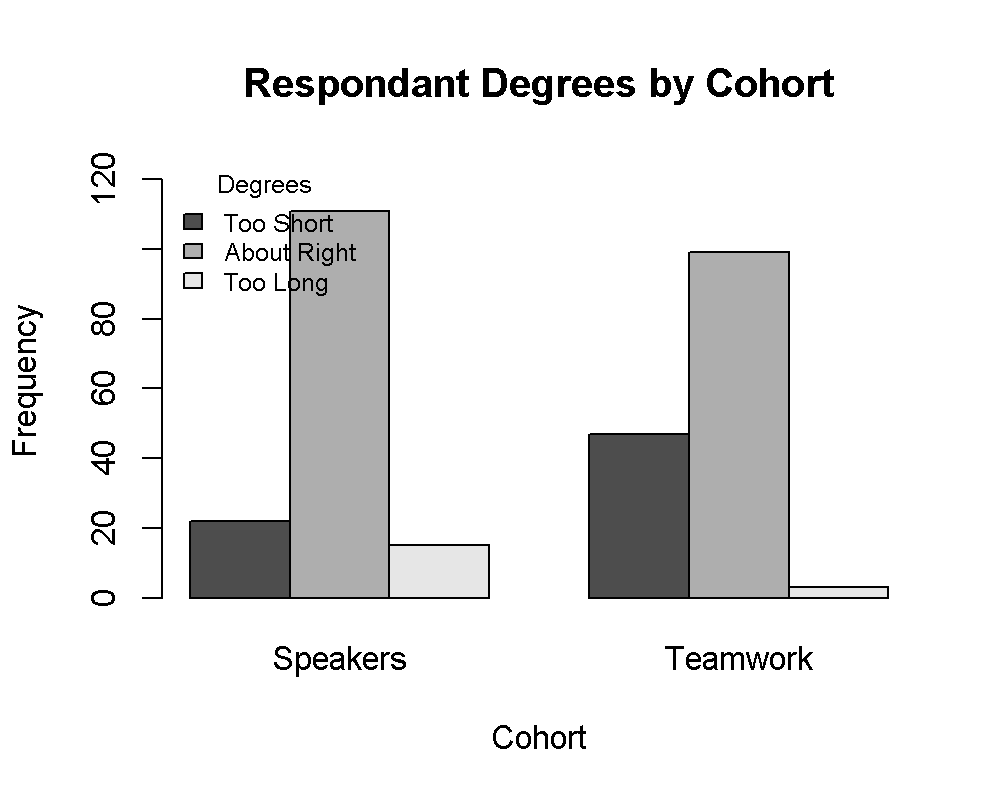
This question was paired with a qualitative open ended questions which asked “Please provide any additional feedback regarding the session videos:” The answers were in line with the scaled responses.

## To preserve the annonymity of participants, the results from this block are hidden.   
TEXT <- select(Qualitative, Cohort, TEXT12)  
TEXT <- filter(TEXT, TEXT12>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

## Time Allocation

* The time allotted each work for speakers/teamwork was:
  + Too Short
  + About Right
  + Too Long

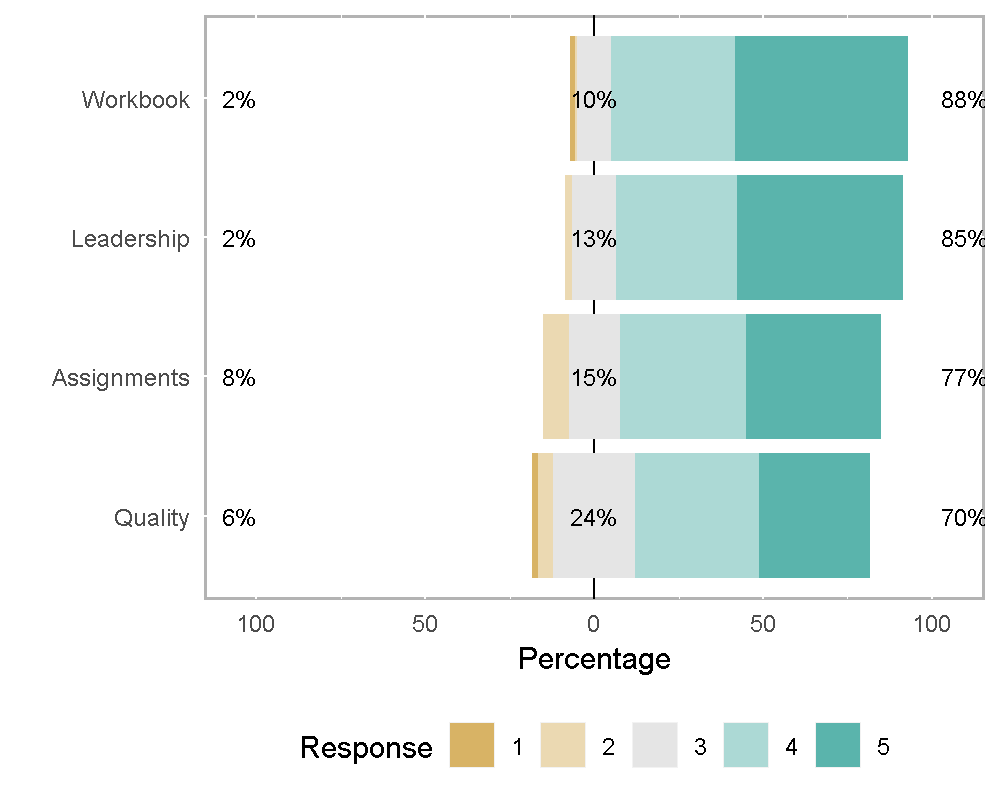
# Extracting data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Speakers = Q13, Teamwork = Q14)  
  
## Combining Q13 & Q14 into one column  
Data <- Data %>%  
 gather(Speakers, Teamwork, key = "Question", value = "Answer")  
  
## Create ordered factors of answer data.  
## Order factor levels, otherwise R will alphabetize them.  
Data$Answer = factor(Data$Answer, ordered = TRUE, levels = TimeL)  
Data$Team = factor(Data$Team, levels=unique(Data$Team))  
Data$Question = factor(Data$Question, levels=unique(Data$Question))  
Data$ID = factor(Data$ID, levels=unique(Data$ID))  
Data$Cohort = factor(Data$Cohort, ordered = TRUE, levels = CohortL)  
  
barplot(table(Data$Answer, Data$Question, useNA = "no"),  
 beside=TRUE,  
 ylim=c(0,125), # adjust to remove legend overlap  
 xlab="Cohort",  
 ylab="Frequency",  
 args.legend=list(cex=0.75, x = "topleft", bty = "n", title="Degrees"),   
 cex.names=1,  
 legend.text=c(levels(Data$Answer)),  
 main = "Respondant Degrees by Cohort")



## General Questions

* Please rate from 1-5 for each category (scale 1=poor, 3=average, 5=excellent)
  + Assignments: How clear were the instructions and assignments?
  + Workbook: How helpful was the Participant Workbook?
  + Leadership: How helpful was the Leadership Team?
  + Quality: How would you rate the quality of the training program overall?

# Extracting data for analysis  
Data <- select(Ordinal, Assignments=Q18\_1, Workbook=Q18\_2, Leadership=Q18\_3, Quality=Q18\_4)  
  
## Create ordered factors of Likert data  
Data$Assignments = factor(Data$Assignments, ordered = TRUE, levels = ScaleL)  
Data$Workbook = factor(Data$Workbook,ordered = TRUE, levels = ScaleL)  
Data$Leadership = factor(Data$Leadership,ordered = TRUE, levels = ScaleL)  
Data$Quality = factor(Data$Quality,ordered = TRUE, levels = ScaleL)  
  
# Calculate Likert Percentages  
Result = likert(Data)  
  
## Bar Ploat  
plot(Result, type="bar")



# Extracting data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Assignments=Q18\_1, Workbook=Q18\_2, Leadership=Q18\_3, Quality=Q18\_4)  
  
# Combining questions into one column  
Data <- Data %>%  
 gather(`Assignments`, `Workbook`, `Leadership`, `Quality`, key = "Question", value = "Likert")  
  
## Create ordered factors of Likert data  
Data$ID = factor(Data$ID, levels=unique(Data$ID), ordered = TRUE)  
Data$Cohort = factor(Data$Cohort, levels=unique(Data$Cohort), ordered = TRUE)  
Data$Team = factor(Data$Team, levels=unique(Data$Team), ordered = TRUE)  
Data$Likert.f = factor(Data$Likert, ordered = TRUE, levels = ScaleL)  
Data$Question = factor(Data$Question, levels=unique(Data$Question))

**Responses Summarized by Question**

# Summary treating Likert data as numeric data  
Summarize(Likert ~ Question,  
 data=Data,  
 digits = 3)

## Question n nvalid mean sd min Q1 median Q3 max  
## 1 Assignments 151 145 4.1 0.92 2 4 4 5 5  
## 2 Workbook 151 147 4.3 0.80 1 4 5 5 5  
## 3 Leadership 151 145 4.3 0.78 2 4 4 5 5  
## 4 Quality 151 115 4.0 0.95 1 3 4 5 5

**Responses Summarized by Cohort**

Summarize(Likert ~ Cohort,  
 data=Data,  
 digits = 3)

## Cohort n nvalid mean sd min Q1 median Q3 max  
## 1 3 168 158 4.2 0.83 2 4 4 5 5  
## 2 2 136 131 4.3 0.82 1 4 4 5 5  
## 3 1 132 95 4.4 0.86 1 4 5 5 5  
## 4 4 168 168 3.9 0.91 1 3 4 5 5

\*\* Analysis of Variance\*\*

# Preparing data for analysis  
Data <- select(Ordinal, ID, Cohort, Team, Degree, Race, Age, Discipline, Employment, Income,  
 Relationship, Q18\_1:Q18\_4)  
Data <- mutate(Data, sum=Q18\_1+Q18\_2+Q18\_3+Q18\_4)  
  
# Checking for statistically significant differences using Anova  
Result <- anova(lm(sum~Cohort+Team+Degree+Age+Race+Discipline+Employment+Income+Relationship,data=Data))  
print(Result)

## Analysis of Variance Table  
##   
## Response: sum  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Cohort 1 37.1 37.1 25.77 0.0023 \*\*  
## Team 12 63.4 5.3 3.67 0.0607 .   
## Degree 3 9.7 3.2 2.24 0.1842   
## Age 4 63.7 15.9 11.06 0.0062 \*\*  
## Race 7 87.5 12.5 8.68 0.0088 \*\*  
## Discipline 7 28.6 4.1 2.83 0.1129   
## Employment 12 105.1 8.8 6.08 0.0184 \*   
## Income 6 45.2 7.5 5.23 0.0321 \*   
## Relationship 5 71.4 14.3 9.91 0.0073 \*\*  
## Residuals 6 8.6 1.4   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Overall Session Feedback

This question was paired with a qualitative open ended questions which asked **“Please provide any additional feedback about the sessions overall?”** The answers were in line with the scaled responses.

## To preserve the annonymity of participants, the results from this block are hidden. }  
TEXT <- select(Qualitative, Cohort, TEXT35)  
TEXT <- filter(TEXT, TEXT35>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

## Learner Time Committment

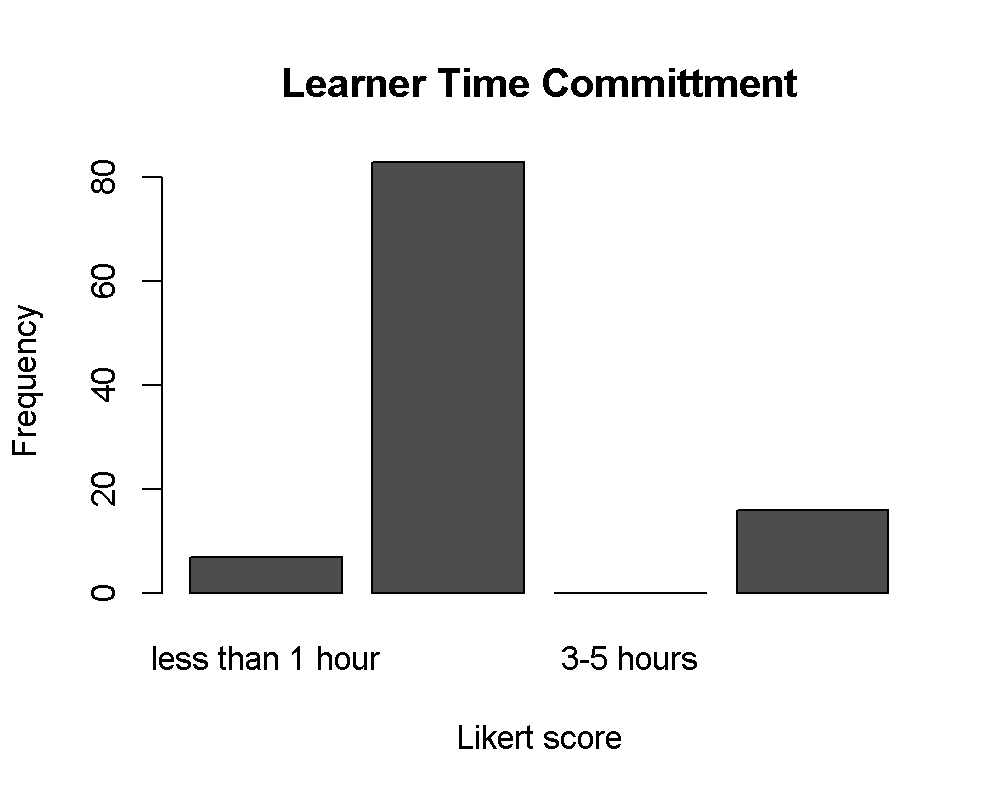
“Q19 - How many hours a week on average did your team meet outside of the program? Would you recommend this program to other women?

**Need to complete this analysis., Fix inconsistencies with scale.**

# Preparing data for analysis  
Data <- select(Ordinal, ID, Cohort, Q19)  
Data$Question = factor("Q19")  
  
# Fix inconsistency in factor levels  
library(stringr)  
Data$Q19 <- str\_replace\_all(Data$Q19, "1-2 hours", "1-2 hrs")  
Data$Q19 <- str\_replace\_all(Data$Q19, "more than 5 hours", "More than 5 hrs")  
Data$Q19 = factor(Data$Q19, ordered = TRUE, levels=c("less than 1 hour", "1-2 hrs", "3-5 hours", "More than 5 hrs"))  
levels(Data$Q19)

## [1] "less than 1 hour" "1-2 hrs" "3-5 hours"   
## [4] "More than 5 hrs"

# Creating the crosstab of Likert data as nominal data  
XT = xtabs(~ Question + Q19, data=Data)  
  
# Creating report output  
  
barplot(XT,  
 col = gray.colors(1),  
 xlab="Likert score",  
 ylab="Frequency",  
 main="Learner Time Committment")



# Overall Program Feedback

This question was paired with a qualitative open ended questions which asked **“What more could the program team do to positively affect perceptions around the barriers to entry?”** The answers were in line with the scaled responses.

## To preserve the annonymity of participants, the results from this block are hidden.   
TEXT <- select(Qualitative, Cohort, TEXT17)  
TEXT <- filter(TEXT, TEXT17>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

This question was paired with a qualitative open ended questions which asked **“Please share any additional information regarding the program and your experience:”** The answers were in line with the scaled responses.

## To preserve the annonymity of participants, the results from this block are hidden.   
TEXT <- select(Qualitative, Cohort, TEXT23)  
TEXT <- filter(TEXT, TEXT23>"")  
TEXT <- arrange(TEXT, desc(Cohort))  
# `r kable(TEXT)`

# Compiling Qualitative Responses

In this analysis, all qualitative data will be compiled and prepared to export to Nvivo for thematic coding and analysis.

# Competency Analysis

These questions are Likert Scale questions (Circle one for each category. Scale 1=poor, 3=average, 5=excellent) which included two side by side scales. One for “Before Ewits”, and one for “After Ewits”.

* Q36: Please rate the following benefits/skills/knowledge levels in terms of how much you feel your participation in the ewitsÂ® program has impacted your personal understanding for each area.
  + Commercialization: The technology commercialization process
  + Research: How to conduct market research
  + Plan: How to write a business plan
  + Presentation: How to develop an investor presentation
  + Funding: Funding sources for startup businesses
* Research identifies various barriers to entry for women in assuming leadership positions in technology startups. Please rate the following in terms of how much you feel your participation in the ewitsÂ® program has impacted your personal perceived challenge(s).
  + Support: Support Systems (mentoring, networking)
  + Confidence: Confidence
  + RoleModels: Role Models
  + Initiative: Self-Initiative (ie: not waiting for an invitation)
  + Balance: Work/Life/Family Balance Issues

# Extracting data for analysis  
Skills <- select(Competency, Commercialization=Q36.2\_1, Research=Q36.2\_2, Plan=Q36.2\_3, Presentation=Q36.2\_4, Funding=Q36.2\_5)  
Barriers <- select(Competency, Support=Q16.2\_1, Confidence=Q16.2\_2, RoleModels=Q16.2\_3, Initiative=Q16.2\_4, Balance=Q16.2\_5)  
SkillsBefore <- select(Competency, Commercialization=Q36.1\_1, Research=Q36.1\_2, Plan=Q36.1\_3, Presentation=Q36.1\_4, Funding=Q36.1\_5)  
BarriersBefore <- select(Competency, Support=Q16.1\_1, Confidence=Q16.1\_2, RoleModels=Q16.1\_3, Initiative=Q16.1\_4, Balance=Q16.1\_5)  
  
## Create ordered factors of Likert data  
Skills$Commercialization = factor(Skills$Commercialization, ordered = TRUE, levels = ScaleL)  
Skills$Research = factor(Skills$Research,ordered = TRUE, levels = ScaleL)  
Skills$Plan = factor(Skills$Plan,ordered = TRUE, levels = ScaleL)  
Skills$Presentation = factor(Skills$Presentation,ordered = TRUE, levels = ScaleL)  
Skills$Funding = factor(Skills$Funding,ordered = TRUE, levels = ScaleL)  
  
Barriers$Support = factor(Barriers$Support, ordered = TRUE, levels = ScaleL)  
Barriers$Confidence = factor(Barriers$Confidence,ordered = TRUE, levels = ScaleL)  
Barriers$RoleModels = factor(Barriers$RoleModels,ordered = TRUE, levels = ScaleL)  
Barriers$Initiative = factor(Barriers$Initiative,ordered = TRUE, levels = ScaleL)  
Barriers$Balance = factor(Barriers$Balance,ordered = TRUE, levels = ScaleL)  
  
SkillsBefore$Commercialization = factor(SkillsBefore$Commercialization, ordered = TRUE, levels = ScaleL)  
SkillsBefore$Research = factor(SkillsBefore$Research,ordered = TRUE, levels = ScaleL)  
SkillsBefore$Plan = factor(SkillsBefore$Plan,ordered = TRUE, levels = ScaleL)  
SkillsBefore$Presentation = factor(SkillsBefore$Presentation,ordered = TRUE, levels = ScaleL)  
SkillsBefore$Funding = factor(SkillsBefore$Funding,ordered = TRUE, levels = ScaleL)  
  
BarriersBefore$Support = factor(BarriersBefore$Support, ordered = TRUE, levels = ScaleL)  
BarriersBefore$Confidence = factor(BarriersBefore$Confidence,ordered = TRUE, levels = ScaleL)  
BarriersBefore$RoleModels = factor(BarriersBefore$RoleModels,ordered = TRUE, levels = ScaleL)  
BarriersBefore$Initiative = factor(BarriersBefore$Initiative,ordered = TRUE, levels = ScaleL)  
BarriersBefore$Balance = factor(BarriersBefore$Balance,ordered = TRUE, levels = ScaleL)  
  
# Calculate Likert Percentages and produce bar plot.  
  
print("Skills Before Ewits")

## [1] "Skills Before Ewits"

Result = likert(SkillsBefore)  
plot(Result, type="bar", main="Competencies & Skills")  
  
print("Skills After Ewits")

## [1] "Skills After Ewits"

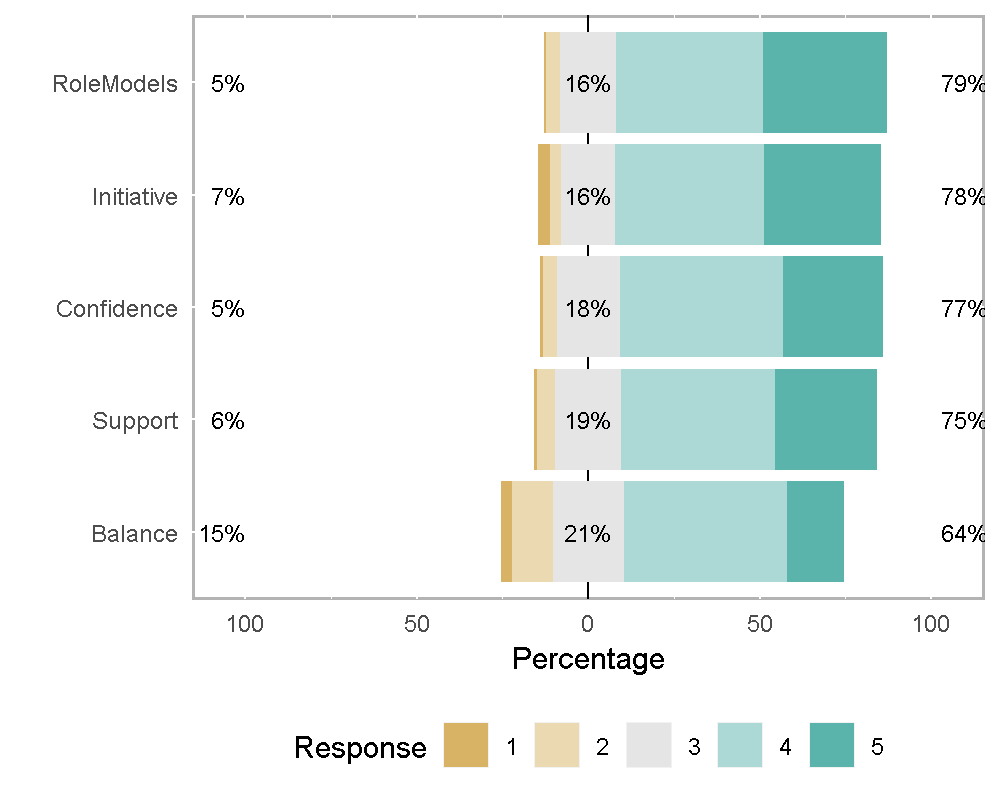
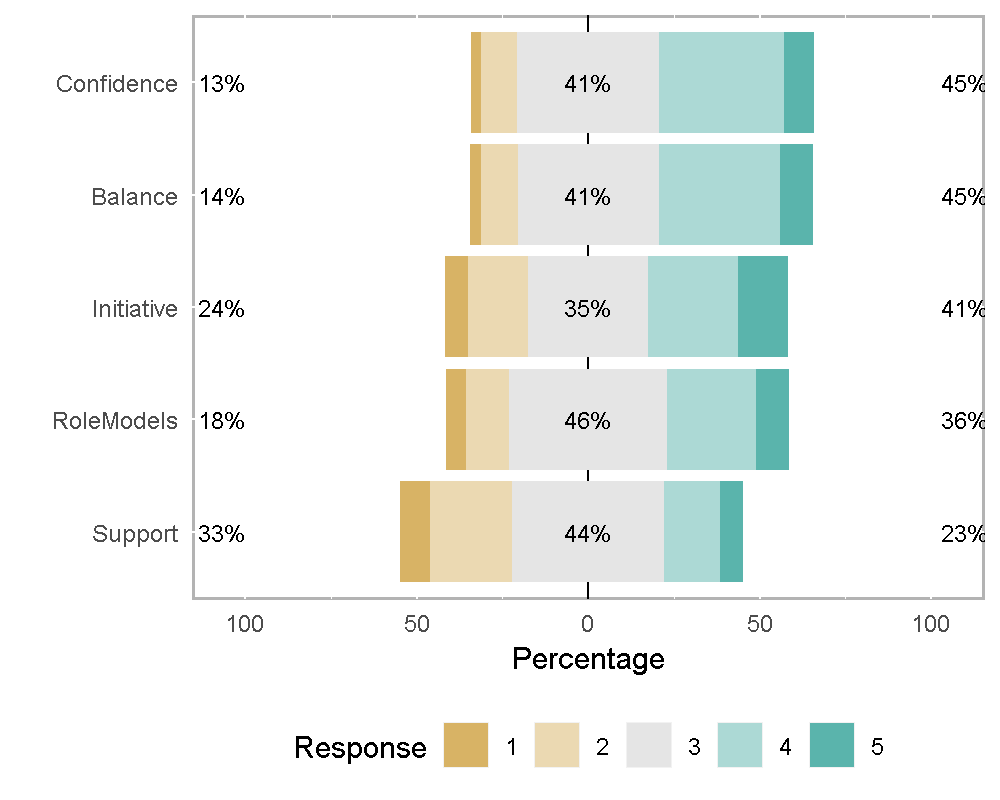
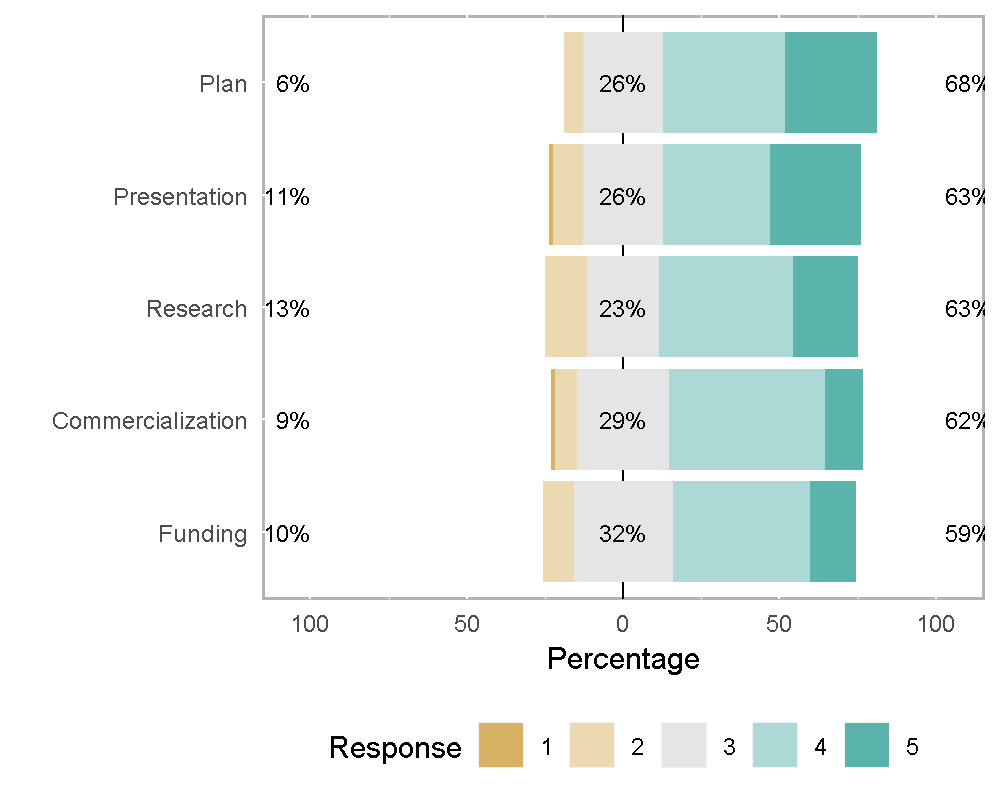
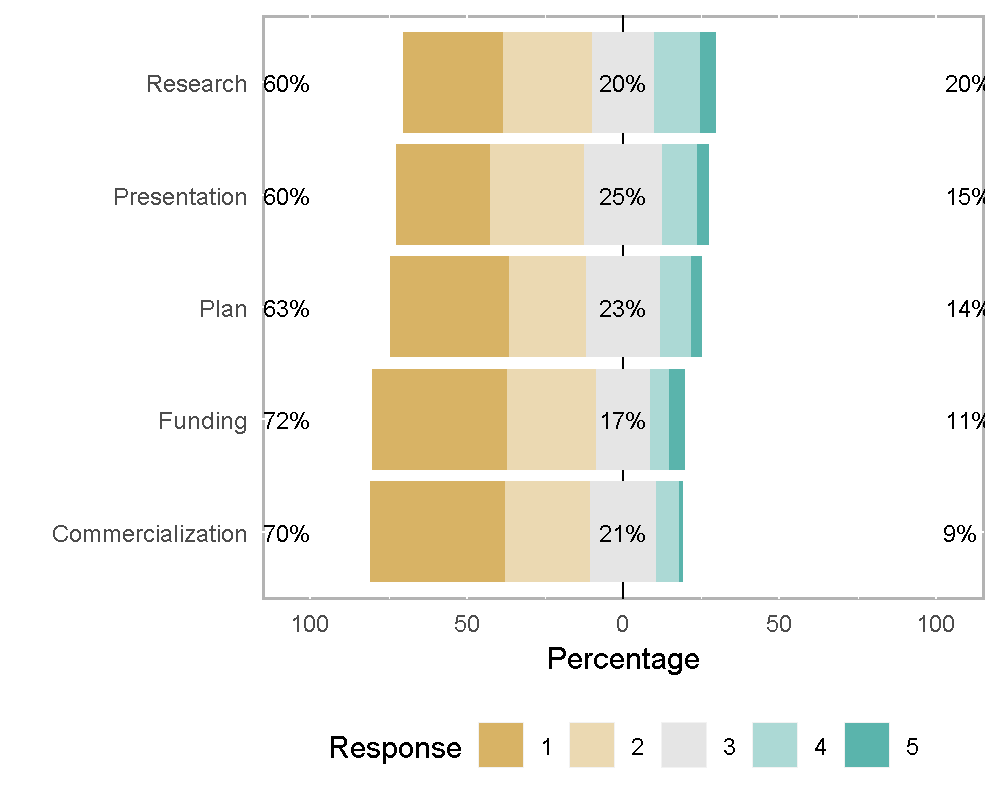
Result = likert(Skills)  
plot(Result, type="bar", main="Competencies & Skills")  
  
print("Bariers Before Ewits")

## [1] "Bariers Before Ewits"

Result = likert(BarriersBefore)  
plot(Result, type="bar", main="Barriers to Participation")  
  
print("Bariers After Ewits")

## [1] "Bariers After Ewits"

Result = likert(Barriers)  
plot(Result, type="bar", main="Barriers to Participation")  
  
## Bar Ploat



# Extracting data for analysis  
Skills <-   
 select(Competency, ID, Cohort, Team, Degree, Commercialization=Q36.2\_1, Research=Q36.2\_2, Plan=Q36.2\_3, Presentation=Q36.2\_4, Funding=Q36.2\_5)  
Barriers <-   
 select(Competency, ID, Cohort, Team, Degree, Support=Q16.2\_1, Confidence=Q16.2\_2, RoleModels=Q16.2\_3, Initiative=Q16.2\_4, Balance=Q16.2\_5)  
SkillsBefore <-   
 select(Competency, ID, Cohort, Team, Degree, Commercialization=Q36.1\_1, Research=Q36.1\_2, Plan=Q36.1\_3, Presentation=Q36.1\_4, Funding=Q36.1\_5)  
BarriersBefore <-   
 select(Competency, ID, Cohort, Team, Degree, Support=Q16.1\_1, Confidence=Q16.1\_2, RoleModels=Q16.1\_3, Initiative=Q16.1\_4, Balance=Q16.1\_5)  
  
# Combining questions into one column  
Skills <- Skills %>%  
 gather(`Commercialization`, `Research`, `Plan`, `Presentation`, `Funding`, key = "Question", value = "After")  
Barriers <- Barriers %>%  
 gather(`Support`, `Confidence`, `RoleModels`, `Initiative`, `Balance`, key = "Question", value = "After")  
SkillsBefore <- SkillsBefore %>%  
 gather(`Commercialization`, `Research`, `Plan`, `Presentation`, `Funding`, key = "Question", value = "Before")  
BarriersBefore <- BarriersBefore %>%  
 gather(`Support`, `Confidence`, `RoleModels`, `Initiative`, `Balance`, key = "Question", value = "Before")  
  
# Join the Before & After data sets  
Skills <- left\_join(SkillsBefore, Skills)

## Joining, by = c("ID", "Cohort", "Team", "Degree", "Question")

Barriers <- left\_join(BarriersBefore, Barriers)

## Joining, by = c("ID", "Cohort", "Team", "Degree", "Question")

# Adding a change column  
Skills <- mutate(Skills, Change=After-Before)  
Barriers <- mutate(Barriers, Change=After-Before)  
  
# Combining Before & After into one column  
Skills <- Skills %>%  
 gather(`Before`, `After`, key = "When", value = "Likert")  
Barriers <- Barriers %>%  
 gather(`Before`, `After`, key = "When", value = "Likert")  
  
  
## Create ordered factors of Likert data  
Skills$ID = factor(Skills$ID, levels=unique(Skills$ID), ordered = TRUE)  
Skills$Cohort = factor(Skills$Cohort, levels = CohortL, ordered = TRUE)  
Skills$Team = factor(Skills$Team, levels=unique(Skills$Team), ordered = TRUE)  
Skills$Degree = factor(Skills$Degree, levels = DegreeL, ordered=TRUE)  
Skills$Likert.f = factor(Skills$Likert, ordered = TRUE, levels = ScaleL)  
Skills$When = factor(Skills$When, levels=unique(Skills$When))  
Skills$Question = factor(Skills$Question, levels=unique(Skills$Question))  
  
Barriers$ID = factor(Barriers$ID, levels=unique(Barriers$ID), ordered = TRUE)  
Barriers$Cohort = factor(Barriers$Cohort, levels = CohortL, ordered = TRUE)  
Barriers$Team = factor(Barriers$Team, levels=unique(Barriers$Team), ordered = TRUE)  
Barriers$Degree = factor(Barriers$Degree, levels = DegreeL, ordered=TRUE)  
Barriers$Likert.f = factor(Barriers$Likert, ordered = TRUE, levels = ScaleL)  
Barriers$When = factor(Barriers$When, levels=unique(Barriers$When))  
Barriers$Question = factor(Barriers$Question, levels=unique(Barriers$Question))  
  
#Verify still dataframe  
Skills <- as.data.frame(Skills)  
Barriers <- as.data.frame(Barriers)

**Responses Summarized by Question**

# Summary treating Likert data as numeric data  
Summarize(Likert ~ When + Question,  
 data=Skills,  
 digits = 3)

## When Question n nvalid mean sd min Q1 median Q3 max  
## 1 Before Commercialization 151 81 2.0 1.03 1 1 2 3 5  
## 2 After Commercialization 151 82 3.6 0.84 1 3 4 4 5  
## 3 Before Research 151 81 2.3 1.21 1 1 2 3 5  
## 4 After Research 151 82 3.7 0.95 2 3 4 4 5  
## 5 Before Plan 151 81 2.2 1.16 1 1 2 3 5  
## 6 After Plan 151 82 3.9 0.89 2 3 4 5 5  
## 7 Before Presentation 151 80 2.3 1.13 1 1 2 3 5  
## 8 After Presentation 151 82 3.8 1.01 1 3 4 5 5  
## 9 Before Funding 151 81 2.0 1.15 1 1 2 3 5  
## 10 After Funding 151 82 3.6 0.85 2 3 4 4 5

Summarize(Likert ~ When + Question,  
 data=Barriers,  
 digits = 3)

## When Question n nvalid mean sd min Q1 median Q3 max  
## 1 Before Support 151 104 2.9 1.01 1 2.0 3 3 5  
## 2 After Support 151 147 4.0 0.88 1 3.5 4 5 5  
## 3 Before Confidence 151 104 3.4 0.89 1 3.0 3 4 5  
## 4 After Confidence 151 147 4.0 0.84 1 4.0 4 5 5  
## 5 Before RoleModels 151 104 3.2 0.98 1 3.0 3 4 5  
## 6 After RoleModels 151 147 4.1 0.86 1 4.0 4 5 5  
## 7 Before Initiative 151 103 3.2 1.12 1 3.0 3 4 5  
## 8 After Initiative 151 147 4.0 0.97 1 4.0 4 5 5  
## 9 Before Balance 151 102 3.4 0.91 1 3.0 3 4 5  
## 10 After Balance 151 145 3.6 1.01 1 3.0 4 4 5

**Analysis of Variance**

# Checking for statistically significant differences using Anova  
Result <- anova(lm(Likert~When+Cohort,data=Skills))  
print(Result)

## Analysis of Variance Table  
##   
## Response: Likert  
## Df Sum Sq Mean Sq F value Pr(>F)   
## When 1 516 516 519.9 < 2e-16 \*\*\*  
## Cohort 1 58 58 58.6 5.6e-14 \*\*\*  
## Residuals 811 806 1   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Result <- anova(lm(Likert~When+Cohort,data=Barriers))  
print(Result)

## Analysis of Variance Table  
##   
## Response: Likert  
## Df Sum Sq Mean Sq F value Pr(>F)   
## When 1 160 159.6 183.7 < 2e-16 \*\*\*  
## Cohort 3 62 20.6 23.7 6.6e-15 \*\*\*  
## Residuals 1245 1081 0.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Cohort 2016 has 40 completers, and 42 surveys.  
S2016 <- read.csv("./data/2016\_Summative.csv", sep = ",", header = TRUE)  
nrow(S2016)  
  
# Rename ID Column  
S2016 <- rename(S2016, ID = Ã¯..V1)  
  
# Rename Cohort column and factor  
S2016 <- rename(S2016, Cohort = Q1)  
S2016$Cohort <- factor(S2016$Cohort, levels=c("2013", "2014", "2015", "2016"), ordered=TRUE)  
levels = c("entirely disagree", "mostly disagree", "somewhat disagree", "neither agree or disagree", "somewhat agree", "mostly agree", "entirely agree")  
  
S2016 <- select(S2016, ID, Cohort, Q50\_1:Q52\_6b)  
  
# Extracting data for analysis  
#Skills <-   
# select(Competency, ID, Cohort, Team, Degree, Commercialization=Q36.2\_1, Research=Q36.2\_2, Plan=Q36.2\_3, Presentation=Q36.2\_4, Funding=Q36.2\_5)  
#Barriers <-   
# select(Competency, ID, Cohort, Team, Degree, Support=Q16.2\_1, Confidence=Q16.2\_2, RoleModels=Q16.2\_3, Initiative=Q16.2\_4, Balance=Q16.2\_5)  
#SkillsBefore <-   
# select(Competency, ID, Cohort, Team, Degree, Commercialization=Q36.1\_1, Research=Q36.1\_2, Plan=Q36.1\_3, Presentation=Q36.1\_4, Funding=Q36.1\_5)  
#BarriersBefore <-   
# select(Competency, ID, Cohort, Team, Degree, Support=Q16.1\_1, Confidence=Q16.1\_2, RoleModels=Q16.1\_3, Initiative=Q16.1\_4, Balance=Q16.1\_5)  
  
# Combining questions into one column  
#Skills <- Skills %>%  
# gather(`Commercialization`, `Research`, `Plan`, `Presentation`, `Funding`, key = "Question", value = "After")  
#Barriers <- Barriers %>%  
# gather(`Support`, `Confidence`, `RoleModels`, `Initiative`, `Balance`, key = "Question", value = "After")  
#SkillsBefore <- SkillsBefore %>%  
# gather(`Commercialization`, `Research`, `Plan`, `Presentation`, `Funding`, key = "Question", value = "Before")  
#BarriersBefore <- BarriersBefore %>%  
# gather(`Support`, `Confidence`, `RoleModels`, `Initiative`, `Balance`, key = "Question", value = "Before")  
  
# Join the Before & After data sets  
#Skills <- left\_join(SkillsBefore, Skills)  
#Barriers <- left\_join(BarriersBefore, Barriers)  
  
# Adding a change column  
#Skills <- mutate(Skills, Change=After-Before)  
#Barriers <- mutate(Barriers, Change=After-Before)  
  
# Combining Before & After into one column  
#Skills <- Skills %>%  
# gather(`Before`, `After`, key = "When", value = "Likert")  
#Barriers <- Barriers %>%  
# gather(`Before`, `After`, key = "When", value = "Likert")  
  
  
## Create ordered factors of Likert data  
#Skills$ID = factor(Skills$ID, levels=unique(Skills$ID), ordered = TRUE)  
#Skills$Cohort = factor(Skills$Cohort, levels = CohortL, ordered = TRUE)  
#Skills$Team = factor(Skills$Team, levels=unique(Skills$Team), ordered = TRUE)  
#Skills$Degree = factor(Skills$Degree, levels = DegreeL, ordered=TRUE)  
#Skills$Likert.f = factor(Skills$Likert, ordered = TRUE, levels = ScaleL)  
#Skills$When = factor(Skills$When, levels=unique(Skills$When))  
#Skills$Question = factor(Skills$Question, levels=unique(Skills$Question))  
#  
#Barriers$ID = factor(Barriers$ID, levels=unique(Barriers$ID), ordered = TRUE)  
#Barriers$Cohort = factor(Barriers$Cohort, levels = CohortL, ordered = TRUE)  
#Barriers$Team = factor(Barriers$Team, levels=unique(Barriers$Team), ordered = TRUE)  
#Barriers$Degree = factor(Barriers$Degree, levels = DegreeL, ordered=TRUE)  
#Barriers$Likert.f = factor(Barriers$Likert, ordered = TRUE, levels = ScaleL)  
#Barriers$When = factor(Barriers$When, levels=unique(Barriers$When))  
#Barriers$Question = factor(Barriers$Question, levels=unique(Barriers$Question))  
  
#Verify still dataframe  
#Skills <- as.data.frame(Skills)  
#Barriers <- as.data.frame(Barriers)

**Responses Summarized by Question**

# Summary treating Likert data as numeric data  
Summarize(Likert ~ When + Question,  
 data=Skills,  
 digits = 3)

## When Question n nvalid mean sd min Q1 median Q3 max  
## 1 Before Commercialization 151 81 2.0 1.03 1 1 2 3 5  
## 2 After Commercialization 151 82 3.6 0.84 1 3 4 4 5  
## 3 Before Research 151 81 2.3 1.21 1 1 2 3 5  
## 4 After Research 151 82 3.7 0.95 2 3 4 4 5  
## 5 Before Plan 151 81 2.2 1.16 1 1 2 3 5  
## 6 After Plan 151 82 3.9 0.89 2 3 4 5 5  
## 7 Before Presentation 151 80 2.3 1.13 1 1 2 3 5  
## 8 After Presentation 151 82 3.8 1.01 1 3 4 5 5  
## 9 Before Funding 151 81 2.0 1.15 1 1 2 3 5  
## 10 After Funding 151 82 3.6 0.85 2 3 4 4 5

Summarize(Likert ~ When + Question,  
 data=Barriers,  
 digits = 3)

## When Question n nvalid mean sd min Q1 median Q3 max  
## 1 Before Support 151 104 2.9 1.01 1 2.0 3 3 5  
## 2 After Support 151 147 4.0 0.88 1 3.5 4 5 5  
## 3 Before Confidence 151 104 3.4 0.89 1 3.0 3 4 5  
## 4 After Confidence 151 147 4.0 0.84 1 4.0 4 5 5  
## 5 Before RoleModels 151 104 3.2 0.98 1 3.0 3 4 5  
## 6 After RoleModels 151 147 4.1 0.86 1 4.0 4 5 5  
## 7 Before Initiative 151 103 3.2 1.12 1 3.0 3 4 5  
## 8 After Initiative 151 147 4.0 0.97 1 4.0 4 5 5  
## 9 Before Balance 151 102 3.4 0.91 1 3.0 3 4 5  
## 10 After Balance 151 145 3.6 1.01 1 3.0 4 4 5

**Analysis of Variance**

# Checking for statistically significant differences using Anova  
Result <- anova(lm(Likert~When+Cohort,data=Skills))  
print(Result)

## Analysis of Variance Table  
##   
## Response: Likert  
## Df Sum Sq Mean Sq F value Pr(>F)   
## When 1 516 516 519.9 < 2e-16 \*\*\*  
## Cohort 1 58 58 58.6 5.6e-14 \*\*\*  
## Residuals 811 806 1   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Result <- anova(lm(Likert~When+Cohort,data=Barriers))  
print(Result)

## Analysis of Variance Table  
##   
## Response: Likert  
## Df Sum Sq Mean Sq F value Pr(>F)   
## When 1 160 159.6 183.7 < 2e-16 \*\*\*  
## Cohort 3 62 20.6 23.7 6.6e-15 \*\*\*  
## Residuals 1245 1081 0.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1