The most important foreign languages for
English-speaking job seekers in global development and humanitarian relief
By Ma. Eliza J. Villarino, June 2016

Breaking into the field of global development and humanitarian relief (or the aid industry) can be tough. Employers often ask candidates to have a few years of experience relevant to the organizations' operations, even for entry-level positions. They also prefer if not require applicants to have foreign language skills, as assignments may entail being deployed overseas.

For English speakers who are serious about joining the global development and humanitarian relief industry, an important question could be which foreign language to invest their time in learning. Knowing the answer can also benefit universities offering courses focused on the sector as this can improve not only their curriculum but more importantly career guidance to students.

Talking with recruiters could help. But perhaps a more definitive method would be to look at job ads – tens of thousands of them.

But where can you find and how do you gather those job ads?

If you Google "international development jobs," you'll see Reliefweb among the top search results. On any given day, it has more than 2,000 open job announcements, volume that's comparable if not better than similar jobs boards. Being owned by the United Nations, it makes all job ads and other information available to the public. Reliefweb also allows users to extract data from its archive, which for job ads date from 2011.

Knowing the demand

The focus here would be on job ads written in English and those that require English and one or more langauges.

To determine the extent of the demand for bilingual or multilingual English speakers means to know the proportion of job ads that require or prefer English-speaking candidates to be fluent or have knowledge in one or a combination of certain foreign languages. In this case, that would French, Spanish, Arabic, Chinese and Russian, which together with English make up the official languages of the United Nations, as well as others that some career advice articles have mentioned, namely Portuguese and "local languages."

Collecting the data

Reliefweb offers an API search of its job ads, which you can run in R using the jsonlite package. The API search string filtered the data based on the month and the year when the job ads were posted, and specified the below fields and subfields. The aim is to use the same extracted data for future data science projects.

- id
- date.created
- title
- body, which contains the job description, including foreign language requirements
- theme.name, or expertise
- experience.name, as expressed by the number of years
- country.name
- career_categories.name, or job functions

- type.name, i.e., whether the position is considered a job, consultancy, internship or volunteer opportuning
- source.name, or the name of the employer
- source.type.name, or the type of organization hiring for the job

Documentation on the Reliefweb API indicates that a user can only extract 1,000 data entries per search. This means iterating the searches by setting the "limit" parameter to 1,000 and the "offset" parameter in intervals of 1,000, starting with 0 for the first search, 1,000 for the second search and so forth.

```
library(jsonlite)

fromJSON("http://api.reliefweb.int/v1/jobs?offset=0&limit=1000&preset=analysis&filter[fie
ld]=date.created&filter[value][from]=2015-05-01T00:00:00%2B00:00&filter[value][to]=2015-
05-
31T00:00:00%2B00:00&&fields[include][]title&fields[include][]=body&fields[include][]=them
e.name&fields[include][]=country.name&fields[include][]=type.name&fields[include][]=exper
ience.name&fields[include][]=career_categories.name&fields[include][]=date.created&fields
[include][]=id&fields[include][]=source.name&fields[include][]=source.type.name")

rwjobs1 <- rwjobsraw1$data$fields
```

Cleaning the data

The extracted data initially had nested lists, which made it impossible to save the data frame into a csv file. Calling the llply function from the plyr package resolved this issue.

```
library(plyr)

rwjobs1$theme <- llply(rwjobs1$theme, unlist)
rwjobs1$type <- llply(rwjobs1$type, unlist)
rwjobs1$experience <- llply(rwjobs1$experience, unlist)
rwjobs1$career_categories <- llply(rwjobs1$career_categories, unlist)
rwjobs1$country <- llply(rwjobs1$country, unlist)
rwjobs1$date <- llply(rwjobs1$date, unlist)
rwjobs1$source <- llply(rwjobs1$source, unlist)
```

That action, however, resulted in the appearance of unnecessary characters in the strings. Using the gsub function removed these characters.

```
rwjobs1$source <- gsub("\\c\\(", "", rwjobs1$source)
rwjobs1$source <- gsub("\"", "", rwjobs1$source)
rwjobs1$source <- gsub("\\)", "", rwjobs1$source)

rwjobs1$theme <- gsub("\\c\\(", "", rwjobs1$theme)
rwjobs1$theme <- gsub("\"", "", rwjobs1$theme)
rwjobs1$theme <- gsub("\\)", "", rwjobs1$theme)

rwjobs1$career_categories <- gsub("\\c\\(", "", rwjobs1$career_categories)
rwjobs1$career_categories <- gsub("\"", "", rwjobs1$career_categories)
rwjobs1$career_categories <- gsub("\\)", "", rwjobs1$career_categories)

rwjobs1$date <- gsub("\\c\\(", "", rwjobs1$date)
rwjobs1$date <- gsub("\"", "", rwjobs1$date)
```

The values under the "source" and "date" columns were separated using the strsplit function and the separated values filled new columns called "organization", "organization_type", "year" and "month".

```
library(stringr)

source_split <- strsplit(rwjobs1$source, split = ",")
select_el <- function(x, index) {x[index]}
org_name <- lapply(source_split, select_el, index = 1)
org_type <- lapply(source_split, select_el, index = 2)
rwjobs1$organization <- as.character(org_name)
rwjobs1$organization_type <- as.character(org_type)
rwjobs1$source <- NULL

date_split <- strsplit(rwjobs1$date, split = "-")
select_el <- function(x, index) {x[index]}
year <- lapply(date_split, select_el, index = 1)
month <- lapply(date_split, select_el, index = 2)
day <- lapply(date_split, select_el, index = 3)
```

The whole process enabled saving the data into a csv file. The rbind function allowed the csv files for all searches to be combined into a dataset with **102,343 unique data entries or job ads posted from March 2011 to June 15, 2016**.

```
rwjAll <- rbind(rwj2011, rwj2012, rwj2013, rwj2014, rwj2015, rwj2016)
rwjAll <- rwjAll[!duplicated(rwjAll), ]
```

With the combined dataset, the type, experience and organization_type columns were further cleaned to clarify the categorical values and replace missing or NA values.

```
library(gdata)

rwjAll$jobType1 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$jobType1 <- str_match(rwjAll$type, "Consultancy")
rwjAll$jobType1 <- ifelse(rwjAll$jobType1=="Consultancy", 1, 0)
rwjAll$jobType1 <- unmatrix(rwjAll$jobType1)

rwjAll$jobType2 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$jobType2 <- str_match(rwjAll$type, "Internship")
rwjAll$jobType2 <- ifelse(rwjAll$jobType2=="Internship", 2, 0)
rwjAll$jobType2 <- unmatrix(rwjAll$jobType2)

rwjAll$jobType3 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$jobType3 <- str_match(rwjAll$type, "Job")
rwjAll$jobType3 <- ifelse(rwjAll$jobType3=="Job", 3, 0)
rwjAll$jobType3 <- unmatrix(rwjAll$jobType3)

rwjAll$jobType4 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$jobType4 <- str_match(rwjAll$type, "Volunteer Opportunity")
rwjAll$jobType4 <- ifelse(rwjAll$jobType4=="Volunteer Opportunity", 4, 0)
rwjAll$jobType4 <- unmatrix(rwjAll$jobType4)

rwjAll$jobTypeAll <- rowSums(rwjAll[, 13:16], na.rm = TRUE)

rwjAll$jobTypeAll[rwjAll$jobTypeAll==1] <- "1 Consultancy"
rwjAll$jobTypeAll[rwjAll$jobTypeAll==2] <- "2 Internship"
rwjAll$jobTypeAll[rwjAll$jobTypeAll==3] <- "3 Job"
rwjAll$jobTypeAll[rwjAll$jobTypeAll==4] <- "4 Volunteer Opportunity"
rwjAll$jobTypeAll[rwjAll$jobTypeAll==0] <- "5 Other"

rwjAll$jobType1 <- NULL
rwjAll$jobType2 <- NULL
rwjAll$jobType3 <- NULL
rwjAll$jobType4 <- NULL
rwjAll$jobType5 <- NULL
```

```
rwjAll$type <- NULL
names(rwjAll)[names(rwjAll)=="jobTypeAll"] <- "job_type"

rwjAll$orgType1 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType1 <- str_match(rwjAll$organization_type, "Academic and Research
Institution")
rwjAll$orgType1 <- ifelse(rwjAll$orgType1=="Academic and Research Institution", 1, 0)
rwjAll$orgType1 <- unmatrix(rwjAll$orgType1)

rwjAll$orgType2 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType2 <- str_match(rwjAll$organization_type, "Inc")
rwjAll$orgType2 <- ifelse(rwjAll$orgType2=="Inc", 2, 0)
rwjAll$orgType2 <- unmatrix(rwjAll$orgType2)

rwjAll$orgType3 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType3 <- str_match(rwjAll$organization_type, "Government")
rwjAll$orgType3 <- ifelse(rwjAll$orgType3=="Government", 3, 0)
rwjAll$orgType3 <- unmatrix(rwjAll$orgType3)

rwjAll$orgType4 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType4 <- str_match(rwjAll$organization_type, "Media")
rwjAll$orgType4 <- ifelse(rwjAll$orgType4=="Media", 4, 0)
rwjAll$orgType4 <- unmatrix(rwjAll$orgType4)

rwjAll$orgType5 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType5 <- str_match(rwjAll$organization_type, "Non-governmental Organization")
rwjAll$orgType5 <- ifelse(rwjAll$orgType5=="Non-governmental Organization", 5, 0)
rwjAll$orgType5 <- unmatrix(rwjAll$orgType5)

rwjAll$orgType6 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType6 <- str_match(rwjAll$organization_type, "Red Cross/Red Crescent Movement")
rwjAll$orgType6 <- ifelse(rwjAll$orgType6=="Red Cross/Red Crescent Movement", 6, 0)
rwjAll$orgType6 <- unmatrix(rwjAll$orgType6)

rwjAll$orgType7 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType7 <- str_match(rwjAll$organization_type, "International Organization")
rwjAll$orgType7 <- ifelse(rwjAll$orgType7=="International Organization", 7, 0)
rwjAll$orgType7 <- unmatrix(rwjAll$orgType7)

rwjAll$orgType8 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$orgType8 <- str_match(rwjAll$organization_type, "Other")
rwjAll$orgType8 <- ifelse(rwjAll$orgType8=="Other", 8, 0)
rwjAll$orgType8 <- unmatrix(rwjAll$orgType8)

rwjAll$orgTypeAll <- rowSums(rwjAll[, 18:25], na.rm = TRUE)

rwjAll$orgTypeAll[rwjAll$orgTypeAll==1] <- "1 Academic and Research Institution"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==2] <- "2 Consultancy"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==3] <- "3 Government"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==4] <- "4 Media"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==5] <- "5 Non-governmental Organization"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==6] <- "6 Red Cross/Red Crescent Movement"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==7] <- "7 International Organization"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==8] <- "8 Other"
rwjAll$orgTypeAll[rwjAll$orgTypeAll==0] <- "8 Other"

rwjAll$orgType1 <- NULL
rwjAll$orgType2 <- NULL
rwjAll$orgType3 <- NULL
rwjAll$orgType4 <- NULL
rwjAll$orgType5 <- NULL
rwjAll$orgType6 <- NULL
rwjAll$orgType7 <- NULL
rwjAll$orgType8 <- NULL

rwjAll$organization_type <- NULL
names(rwjAll)[names(rwjAll)=="orgTypeAll"] <- "organization_type"
```

```
rwjAllTest <- as.data.frame(rwjAll)
rwjAllTest$experience <- as.character(rwjAllTest$experience)
rwjAllTest$experience[rwjAllTest$experience=="NULL"] <- "Other"
rwjAllTest$experience[rwjAllTest$experience=="N/A"] <- "Other"
```

Transforming and preparing the data for analysis

Calling the str_match function and ifelse statements, binary columns were added to the dataset to determine the frequencies and later the proportion of job ads requiring or preferring English speakers with foreign language skills.

The following keyword searches of the "body" column formed the basis of the values in the binary columns:

- "is " which denotes that the ad is written English.
- "anglais" which denotes that the ad is written French and seeks English speakers.
- "ingles" which denotes that the ad is written Spanish and seeks English speakers.
- "French" which denotes that the ad asks for applicants to be fluent in or knowledge of French.
- "Spanish" which denotes that the ad asks for applicants to be fluent in or knowledge of Spanish.
- "Arabic" which denotes that the ad asks for applicants to be fluent in or knowledge of Arabic.
- "Chinese" which denotes that the ad asks for applicants to be fluent in or knowledge of Chinese.
- "Russian" which denotes that the ad asks for applicants to be fluent in or knowledge of Russian.
- "Portuguese" which denotes that the ad asks for applicants to be fluent in or knowledge of Portuguese.
- "local languages" which denotes that the ad asks for applicants to be fluent in or knowledge of local languages.

Another binary column was created to combine values for all languages, and this was used to initially filter the dataset.

```
rwjAll$English1 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English1 <- str_match(rwjAll$body, "is ")
rwjAll$English1 <- ifelse(rwjAll$English1 == "is ", 1, 0)
rwjAll$English1 <- unmatrix(rwjAll$English1)

rwjAll$English2 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English2 <- str_match(rwjAll$body, "anglais")
rwjAll$English2 <- ifelse(rwjAll$English2 == "anglais", 1, 0)
rwjAll$English2 <- unmatrix(rwjAll$English2)

rwjAll$English3 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English3 <- str_match(rwjAll$body, "ingles")
rwjAll$English3 <- ifelse(rwjAll$English3 == "ingles", 1, 0)
rwjAll$English3 <- unmatrix(rwjAll$English3)

rwjAll$English_total <- rowSums(rwjAll[, 13:15], na.rm = TRUE)
rwjAll$English_total[rwjAll$English_total == 2] <- 1

rwjAll$French <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$French <- str_match(rwjAll$body, "French")
rwjAll$French <- ifelse(rwjAll$French == "French", 1, 0)
rwjAll$French <- unmatrix(rwjAll$French)

rwjAll$Arabic <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$Arabic <- str_match(rwjAll$body, "Arabic")
rwjAll$Arabic <- ifelse(rwjAll$Arabic=="Arabic", 1, 0)
rwjAll$Arabic <- unmatrix(rwjAll$Arabic)
```

```
rwjAll$Spanish <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$Spanish <- str_match(rwjAll$body, "Spanish")
rwjAll$Spanish <- ifelse(rwjAll$Spanish=="Spanish", 1, 0)
rwjAll$Spanish <- unmatrix(rwjAll$Spanish)

rwjAll$Russian <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$Russian <- str_match(rwjAll$body, "Russian")
rwjAll$Russian <- ifelse(rwjAll$Russian=="Russian", 1, 0)
rwjAll$Russian <- unmatrix(rwjAll$Russian)

rwjAll$Chinese <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$Chinese <- str_match(rwjAll$body, "Chinese")
rwjAll$Chinese <- ifelse(rwjAll$Chinese=="Chinese", 1, 0)
rwjAll$Chinese <- unmatrix(rwjAll$Chinese)

rwjAll$Portuguese <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$Portuguese <- str_match(rwjAll$body, "Portuguese")
rwjAll$Portuguese <- ifelse(rwjAll$Portuguese=="Portuguese", 1, 0)
rwjAll$Portuguese <- unmatrix(rwjAll$Portuguese)

rwjAll$local <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$local <- str_match(rwjAll$body, "local language")
rwjAll$local <- ifelse(rwjAll$local=="local language", 1, 0)
rwjAll$local <- unmatrix(rwjAll$local)

rwjAll$French_Arabic <- rowSums(rwjAll[, c("French", "English2", "Arabic")], na.rm =
TRUE)
rwjAll$French_Arabic[rwjAll$French_Arabic==2] <- 1
rwjAll$French_Arabic[rwjAll$French_Arabic==3] <- 1

rwjAll$French_Spanish <- rowSums(rwjAll[, c("French", "English2", "Spanish",
"English3")], na.rm = TRUE)
rwjAll$French_Spanish[rwjAll$French_Spanish==2] <- 1
rwjAll$French_Spanish[rwjAll$French_Spanish==3] <- 1

rwjAll$French_Arabic_Spanish <- rowSums(rwjAll[, c("French", "English2", "Arabic",
"Spanish", "English3")], na.rm = TRUE)
rwjAll$French_Arabic_Spanish[rwjAll$French_Arabic_Spanish==2] <- 1
rwjAll$French_Arabic_Spanish[rwjAll$French_Arabic_Spanish==3] <- 1
rwjAll$French_Arabic_Spanish[rwjAll$French_Arabic_Spanish==4] <- 1

rwjAll$language_all <- rowSums(rwjAll[, 16:26], na.rm = TRUE)
rwjAll$language_all[rwjAll$language_all == 2] <- 1
rwjAll$language_all[rwjAll$language_all == 3] <- 1
rwjAll$language_all[rwjAll$language_all == 4] <- 1
rwjAll$language_all[rwjAll$language_all == 5] <- 1
rwjAll$language_all[rwjAll$language_all == 6] <- 1
rwjAll$language_all[rwjAll$language_all == 7] <- 1
rwjAll$language_all[rwjAll$language_all == 8] <- 1
rwjAll$language_all[rwjAll$language_all == 9] <- 1
rwjAll$language_all[rwjAll$language_all == 10] <- 1
rwjAll$language_all[rwjAll$language_all == 11] <- 1

library(dplyr)
library(tidyr)

rwjAll_bil <- filter(rwjAll, language_all == 1)
rwjAll_bil <- rwjAll_bil[!duplicated(rwjAll_bil), ]
rwjAll_bil$EF <- rwjAll_bil$English1 + rwjAll_bil$English2 + rwjAll_bil$French
rwjAll_bil$EF[rwjAll_bil$EF==2] <- 1
rwjAll_bil$EF[rwjAll_bil$EF==3] <- 1
rwjAll_bil$EF[rwjAll_bil$EF==4] <- 1
rwjAll_bil$EA <- rwjAll_bil$English1 + rwjAll_bil$Arabic
rwjAll_bil$EA[rwjAll_bil$EA==2] <- 1
rwjAll_bil$ES <- rwjAll_bil$English1 + rwjAll_bil$English3 + rwjAll_bil$Spanish
rwjAll_bil$ES[rwjAll_bil$ES==2] <- 1
rwjAll_bil$ES[rwjAll_bil$ES==3] <- 1
```

```
rwjAll_bil$ER <- rwjAll_bil$English1 + rwjAll_bil$Russian
rwjAll_bil$EC <- rwjAll_bil$English1 + rwjAll_bil$Chinese
rwjAll_bil$EP <- rwjAll_bil$English1 + rwjAll_bil$Portuguese
rwjAll_bil$EL <- rwjAll_bil$English1 + rwjAll_bil$locallibrary(tidyr)
```

The filtered dataset (with 99,310 entries) was filtered anew, by year to compare annual trends.

```
rwjAll_bil16 <- filter(rwjAll_bil, year == 2016)
rwjAll_bil15 <- filter(rwjAll_bil, year == 2015)
rwjAll_bil14 <- filter(rwjAll_bil, year == 2014)
rwjAll_bil13 <- filter(rwjAll_bil, year == 2013)
rwjAll_bil12 <- filter(rwjAll_bil, year == 2012)
rwjAll_bil11 <- filter(rwjAll_bil, year == 2011)
```

To get the proportion of job ads asking for English speakers with certain foreign language skills, the values under the columns showing the frequencies were summed up, excluding the NA values, and then divided by the number of the dataset rows.

```
rwjAll_bil16 <- filter(rwjAll_bil, year == 2016)
total_EF_16 <- sum(rwjAll_bil16$French, na.rm = TRUE) +  sum(rwjAll_bil16$English2, na.rm =
TRUE)
EF_percent_16 <- total_EF_16 / nrow(rwjAll_bil16)
total_EA_16 <- sum(rwjAll_bil16$Arabic, na.rm = TRUE)
EA_percent_16 <- total_EA_16 / nrow(rwjAll_bil16)
total_ES_16 <- sum(rwjAll_bil16$Spanish, na.rm = TRUE) + sum(rwjAll_bil16$English3, na.rm =
TRUE)
ES_percent_16 <- total_ES_16 / nrow(rwjAll_bil16)
total_ER_16 <- sum(rwjAll_bil16$Russian, na.rm = TRUE)
ER_percent_16 <- total_ER_16 / nrow(rwjAll_bil16)
total_EC_16 <- sum(rwjAll_bil16$Chinese, na.rm = TRUE)
EC_percent_16 <- total_EC_16 / nrow(rwjAll_bil16)
total_EP_16 <- sum(rwjAll_bil16$Portuguese, na.rm = TRUE)
EP_percent_16 <- total_EP_16 / nrow(rwjAll_bil16)
total_EL_16 <- sum(rwjAll_bil16$local, na.rm = TRUE)
EL_percent_16 <- total_EL_16 / nrow(rwjAll_bil16)
total_EFA_16 <- sum(rwjAll_bil16$French_Arabic, na.rm = TRUE)
EFA_percent_16 <- total_EFA_16 / nrow(rwjAll_bil16)
total_EFS_16 <- sum(rwjAll_bil16$French_Spanish, na.rm = TRUE)
EFS_percent_16 <- total_EFS_16 / nrow(rwjAll_bil16)
total_EFAS_16 <- sum(rwjAll_bil16$French_Arabic_Spanish, na.rm = TRUE)
EFAS_percent_16 <- total_EFAS_16 / nrow(rwjAll_bil16)

rwjAll_bil15 <- filter(rwjAll_bil, year == 2015)
total_EF_15 <- sum(rwjAll_bil15$French, na.rm = TRUE) + sum(rwjAll_bil15$English2, na.rm =
TRUE)
EF_percent_15 <- total_EF_15 / nrow(rwjAll_bil15)
total_EA_15 <- sum(rwjAll_bil15$Arabic, na.rm = TRUE)
EA_percent_15 <- total_EA_15 / nrow(rwjAll_bil15)
total_ES_15 <- sum(rwjAll_bil15$Spanish, na.rm = TRUE) + sum(rwjAll_bil15$English3, na.rm =
TRUE)
ES_percent_15 <- total_ES_15 / nrow(rwjAll_bil15)
total_ER_15 <- sum(rwjAll_bil15$Russian, na.rm = TRUE)
ER_percent_15 <- total_ER_15 / nrow(rwjAll_bil15)
total_EC_15 <- sum(rwjAll_bil15$Chinese, na.rm = TRUE)
EC_percent_15 <- total_EC_15 / nrow(rwjAll_bil15)
total_EP_15 <- sum(rwjAll_bil15$Portuguese, na.rm = TRUE)
EP_percent_15 <- total_EP_15 / nrow(rwjAll_bil15)
total_EL_15 <- sum(rwjAll_bil15$local, na.rm = TRUE)
EL_percent_15 <- total_EL_15 / nrow(rwjAll_bil15)
total_EFA_15 <- sum(rwjAll_bil15$French_Arabic, na.rm = TRUE)
EFA_percent_15 <- total_EFA_15 / nrow(rwjAll_bil15)
total_EFS_15 <- sum(rwjAll_bil15$French_Spanish, na.rm = TRUE)
EFS_percent_15 <- total_EFS_15 / nrow(rwjAll_bil15)
total_EFAS_15 <- sum(rwjAll_bil15$French_Arabic_Spanish, na.rm = TRUE)
```

```
EFAS_percent_15 <- total_EFAS_15 / nrow(rwjAll_bil15)

rwjAll_bil14 <- filter(rwjAll_bil, year == 2014)
total_EF_14 <- sum(rwjAll_bil14$French, na.rm = TRUE) + sum(rwjAll_bil14$Englis2, na.rm =
TRUE)
EF_percent_14 <- total_EF_14 / nrow(rwjAll_bil14)
total_EA_14 <- sum(rwjAll_bil14$Arabic, na.rm = TRUE)
EA_percent_14 <- total_EA_14 / nrow(rwjAll_bil14)
total_ES_14 <- sum(rwjAll_bil14$Spanish, na.rm = TRUE) +sum(rwjAll_bil14$English3, na.rm =
TRUE)
ES_percent_14 <- total_ES_14 / nrow(rwjAll_bil14)
total_ER_14 <- sum(rwjAll_bil14$Russian, na.rm = TRUE)
ER_percent_14 <- total_ER_14 / nrow(rwjAll_bil14)
total_EC_14 <- sum(rwjAll_bil14$Chinese, na.rm = TRUE)
EC_percent_14 <- total_EC_14 / nrow(rwjAll_bil14)
total_EP_14 <- sum(rwjAll_bil14$Portuguese, na.rm = TRUE)
EP_percent_14 <- total_EP_14 / nrow(rwjAll_bil14)
total_EL_14 <- sum(rwjAll_bil14$local, na.rm = TRUE)
EL_percent_14 <- total_EL_14 / nrow(rwjAll_bil14)
total_EFA_14 <- sum(rwjAll_bil14$French_Arabic, na.rm = TRUE)
EFA_percent_14 <- total_EFA_14 / nrow(rwjAll_bil14)
total_EFS_14 <- sum(rwjAll_bil14$French_Spanish, na.rm = TRUE)
EFS_percent_14 <- total_EFS_14 / nrow(rwjAll_bil14)
total_EFAS_14 <- sum(rwjAll_bil14$French_Arabic_Spanish, na.rm = TRUE)
EFAS_percent_14 <- total_EFAS_14 / nrow(rwjAll_bil14)

rwjAll_bil13 <- filter(rwjAll_bil, year == 2013)
total_EF_13 <- sum(rwjAll_bil13$French, na.rm = TRUE) + sum(rwjAll_bil13$English2, na.rm =
TRUE)
EF_percent_13 <- total_EF_13 / nrow(rwjAll_bil13)
total_EA_13 <- sum(rwjAll_bil13$Arabic, na.rm = TRUE)
EA_percent_13 <- total_EA_13 / nrow(rwjAll_bil13)
total_ES_13 <- sum(rwjAll_bil13$Spanish, na.rm = TRUE) + sum(rwjAll_bil13$English3, na.rm =
TRUE)
ES_percent_13 <- total_ES_13 / nrow(rwjAll_bil13)
total_ER_13 <- sum(rwjAll_bil13$Russian, na.rm = TRUE)
ER_percent_13 <- total_ER_13 / nrow(rwjAll_bil13)
total_EC_13 <- sum(rwjAll_bil13$Chinese, na.rm = TRUE)
EC_percent_13 <- total_EC_13 / nrow(rwjAll_bil13)
total_EP_13 <- sum(rwjAll_bil13$Portuguese, na.rm = TRUE)
EP_percent_13 <- total_EP_13 / nrow(rwjAll_bil13)
total_EL_13 <- sum(rwjAll_bil13$local, na.rm = TRUE)
EL_percent_13 <- total_EL_13 / nrow(rwjAll_bil13)
total_EFA_13 <- sum(rwjAll_bil13$French_Arabic, na.rm = TRUE)
EFA_percent_13 <- total_EFA_13 / nrow(rwjAll_bil13)
total_EFS_13 <- sum(rwjAll_bil13$French_Spanish, na.rm = TRUE)
EFS_percent_13 <- total_EFS_13 / nrow(rwjAll_bil13)
total_EFAS_13 <- sum(rwjAll_bil13$French_Arabic_Spanish, na.rm = TRUE)
EFAS_percent_13 <- total_EFAS_13 / nrow(rwjAll_bil13)

rwjAll_bil12 <- filter(rwjAll_bil, year == 2012)
total_EF_12 <- sum(rwjAll_bil12$French, na.rm = TRUE) + sum(rwjAll_bil12$English2, na.rm =
TRUE)
EF_percent_12 <- total_EF_12 / nrow(rwjAll_bil12)
total_EA_12 <- sum(rwjAll_bil12$Arabic, na.rm = TRUE)
EA_percent_12 <- total_EA_12 / nrow(rwjAll_bil12)
total_ES_12 <- sum(rwjAll_bil12$Spanish, na.rm = TRUE) + sum(rwjAll_bil12$English3, na.rm =
TRUE)
ES_percent_12 <- total_ES_12 / nrow(rwjAll_bil12)
total_ER_12 <- sum(rwjAll_bil12$Russian, na.rm = TRUE)
ER_percent_12 <- total_ER_12 / nrow(rwjAll_bil12)
total_EC_12 <- sum(rwjAll_bil12$Chinese, na.rm = TRUE)
EC_percent_12 <- total_EC_12 / nrow(rwjAll_bil12)
total_EP_12 <- sum(rwjAll_bil12$Portuguese, na.rm = TRUE)
EP_percent_12 <- total_EP_12 / nrow(rwjAll_bil12)
total_EL_12 <- sum(rwjAll_bil12$local, na.rm = TRUE)
EL_percent_12 <- total_EL_12 / nrow(rwjAll_bil12)
total_EFA_12 <- sum(rwjAll_bil12$French_Arabic, na.rm = TRUE)
```

```
EFA_percent_12 <- total_EFA_12 / nrow(rwjAll_bil12)
total_EFS_12 <- sum(rwjAll_bil12$French_Spanish, na.rm = TRUE)
EFS_percent_12 <- total_EFS_12 / nrow(rwjAll_bil12)
total_EFAS_12 <- sum(rwjAll_bil12$French_Arabic_Spanish, na.rm = TRUE)
EFAS_percent_12 <- total_EFAS_12 / nrow(rwjAll_bil12)

rwjAll_bil11 <- filter(rwjAll_bil, year == 2011)
total_EF_11 <- sum(rwjAll_bil11$French, na.rm = TRUE) + sum(rwjAll_bil11$English2, na.rm =
TRUE)
EF_percent_11 <- total_EF_11 / nrow(rwjAll_bil11)
total_EA_11 <- sum(rwjAll_bil11$Arabic, na.rm = TRUE)
EA_percent_11 <- total_EA_11 / nrow(rwjAll_bil11)
total_ES_11 <- sum(rwjAll_bil11$Spanish, na.rm = TRUE) + sum(rwjAll_bil11$English3, na.rm =
TRUE)
ES_percent_11 <- total_ES_11 / nrow(rwjAll_bil11)
total_ER_11 <- sum(rwjAll_bil11$Russian, na.rm = TRUE)
ER_percent_11 <- total_ER_11 / nrow(rwjAll_bil11)
total_EC_11 <- sum(rwjAll_bil11$Chinese, na.rm = TRUE)
EC_percent_11 <- total_EC_11 / nrow(rwjAll_bil11)
total_EP_11 <- sum(rwjAll_bil11$Portuguese, na.rm = TRUE)
EP_percent_11 <- total_EP_11 / nrow(rwjAll_bil11)
total_EL_11 <- sum(rwjAll_bil11$local, na.rm = TRUE)
EL_percent_11 <- total_EL_11 / nrow(rwjAll_bil11)
total_EFA_11 <- sum(rwjAll_bil11$French_Arabic, na.rm = TRUE)
EFA_percent_11 <- total_EFA_11 / nrow(rwjAll_bil11)
total_EFS_11 <- sum(rwjAll_bil11$French_Spanish, na.rm = TRUE)
EFS_percent_11 <- total_EFS_11 / nrow(rwjAll_bil11)
total_EFAS_11 <- sum(rwjAll_bil11$French_Arabic_Spanish, na.rm = TRUE)
EFAS_percent_11 <- total_EFAS_11 / nrow(rwjAll_bil11)
```

The proportion of job ads requiring bilingual and multilingual skills (transformed into whole numbers and rounded to two digits) populated a new dataset that also includes the covered years (2011-2016). This dataset was used for the analysis.

```
year <- c(2011, 2012, 2013, 2014, 2015, 2016)
English_French <- c(EF_percent_11, EF_percent_12, EF_percent_13, EF_percent_14,
EF_percent_15, EF_percent_16)
English_Arabic <- c(EA_percent_11, EA_percent_12, EA_percent_13, EA_percent_14,
EA_percent_15, EA_percent_16)
English_Spanish <- c(ES_percent_11, ES_percent_12, ES_percent_13, ES_percent_14,
ES_percent_15, ES_percent_16)
English_Russian <- c(ER_percent_11, ER_percent_12, ER_percent_13, ER_percent_14,
ER_percent_15, ER_percent_16)
English_Chinese <- c(EC_percent_11, EC_percent_12, EC_percent_13, EC_percent_14,
EC_percent_15, EC_percent_16)
English_Portuguese <- c(EP_percent_11, EP_percent_12, EP_percent_13, EP_percent_14,
EP_percent_15, EP_percent_16)
English_local <- c(EL_percent_11, EL_percent_12, EL_percent_13, EL_percent_14,
EL_percent_15, EL_percent_16)
English_French_Arabic <- c(EFA_percent_11, EFA_percent_12, EFA_percent_13,
EFA_percent_14, EFA_percent_15, EFA_percent_16)
English_French_Spanish <- c(EFS_percent_11, EFS_percent_12, EFS_percent_13,
EFS_percent_14, EFS_percent_15, EFS_percent_16)
English_French_Arabic_Spanish <- c(EFAS_percent_11, EFAS_percent_12, EFAS_percent_13,
EFAS_percent_14, EFAS_percent_15, EFAS_percent_16)

bi_lang_pct <- cbind(year, English_French, English_Arabic, English_Spanish,
English_Russian, English_Chinese, English_Portuguese, English_local,
English_French_Arabic, English_French_Spanish, English_French_Arabic_Spanish)
bi_lang_pct <- as.data.frame(bi_lang_pct)
bi_lang_pct$English_French <- bi_lang_pct$English_French * 100
bi_lang_pct$English_French <- round(bi_lang_pct$English_French, digits = 2)
bi_lang_pct$English_Arabic <- bi_lang_pct$English_Arabic * 100
bi_lang_pct$English_Arabic <- round(bi_lang_pct$English_Arabic, digits = 2)
bi_lang_pct$English_Spanish <- bi_lang_pct$English_Spanish * 100
bi_lang_pct$English_Spanish <- round(bi_lang_pct$English_Spanish, digits = 2)
```

```
bi_lang_pct$English_Russian <- bi_lang_pct$English_Russian * 100
bi_lang_pct$English_Russian <- round(bi_lang_pct$English_Russian, digits = 2)
bi_lang_pct$English_Chinese <- bi_lang_pct$English_Chinese * 100
bi_lang_pct$English_Chinese <- round(bi_lang_pct$English_Chinese, digits = 2)
bi_lang_pct$English_Portuguese <- bi_lang_pct$English_Portuguese * 100
bi_lang_pct$English_Portuguese <- round(bi_lang_pct$English_Portuguese, digits = 2)
bi_lang_pct$English_local <- bi_lang_pct$English_local * 100
bi_lang_pct$English_local <- round(bi_lang_pct$English_local, digits = 2)
bi_lang_pct$English_French_Arabic <- bi_lang_pct$English_French_Arabic * 100
bi_lang_pct$English_French_Arabic <- round(bi_lang_pct$English_French_Arabic, digits = 2)
bi_lang_pct$English_French_Spanish <- bi_lang_pct$English_French_Spanish * 100
bi_lang_pct$English_French_Spanish <- round(bi_lang_pct$English_French_Spanish, digits =
2)
bi_lang_pct$English_French_Arabic_Spanish <- bi_lang_pct$English_French_Arabic_Spanish *
100
bi_lang_pct$English_French_Arabic_Spanish <-
round(bi_lang_pct$English_French_Arabic_Spanish, digits = 2)
```

Analyzing the data

The table below illustrates the dataset with the proportions of job ads seeking bilingual and multilingual speakers.

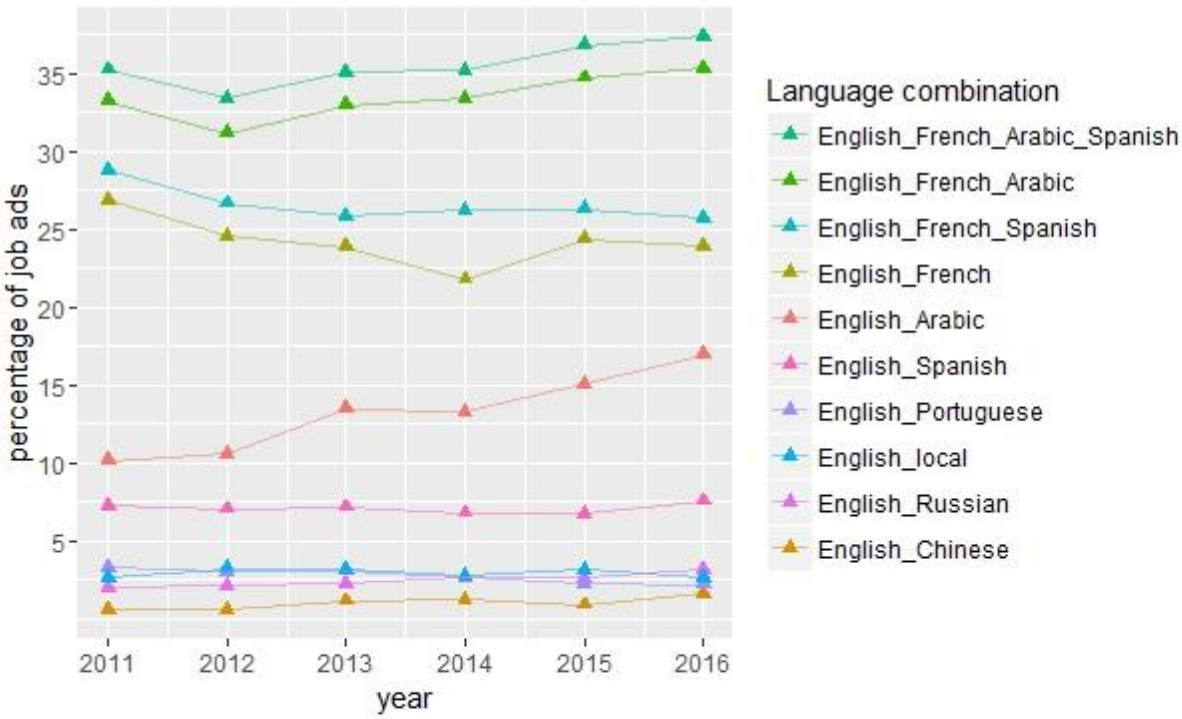
year	English_French	English_Arabic	English_Spanish	English_Russian	English_Chinese	English_Portuguese	English_local	English_French_Arabic	English_French_Spanish	English_French_Arabic_Spanish
2011	26.85	10.12	7.22	1.94	0.59	3.26	2.6	33.22	28.79	35.24
2012	24.53	10.56	7.02	2.06	0.52	3.02	3.21	31.16	26.65	33.42
2013	23.86	13.46	7.16	2.25	1.13	3.04	3.1	32.96	25.82	35.03
2014	21.76	13.26	6.75	2.58	1.23	2.63	2.74	33.37	26.22	35.18
2015	24.37	15.05	6.71	2.58	0.88	2.23	3.11	34.72	26.29	36.81
2016	23.89	16.93	7.54	3.15	1.58	2.18	2.64	35.35	25.67	37.37

Reshaping the data frame allowed for plotting of the results using the ggplot2 package

```
library(ggplot2)

ml_comb <- bi_lang_pct %>% gather(bilingual_combination, percentage, 2:11)

plot1 <- ggplot(ml_comb, aes(x = year, y = percentage, color = bilingual_combination)) +
        geom_point(shape = 17, size = 2) +
        geom_line(size = 0.5, alpha = 0.5) +
        scale_y_continuous(name="percentage of job ads", breaks = c(5, 10, 15, 20, 25,
                           30, 35, 40, 45, 50, 55, 60, 65)) +
        scale_color_discrete(name = "Language combination",
                           breaks = c("English_French_Arabic_Spanish",
                           "English_French_Arabic", "English_French_Spanish",
                           "English_French", "English_Arabic", "English_Spanish",
                           "English_Portuguese", "English_local", "English_Russian",
                           "English_Chinese"))
```



The table and the plot show that:

- English speakers who are fluent in or have knowledge of French have the highest demand among bilingual English job seekers in global development and humanitarian relief. Nearly 24 in 100job ads seek such candidates in 2016, while about 17 in 100 positions requires or prefers English-Arabic speakers. Meanwhile, only more than 1 in 100 assignments need or favor English-Chinese speakers so far in 2016.

- Knowing both French and Arabic allows qualified English speakers to apply for or gives them an edge in nearly 35 in 100 jobs in the aid industry.

- The demand for English-French speakers in global development has declined over the years, from 26.9 percent in 2011 to 23.9 percent as of June 15, 2016. On the other hand, the demand for English-Arabic job candidates is on an upswing, rising from 10.1 percent in 2011 to nearly 16.9 percent by mid-June 2016.

- There's a relatively steady demand for English speakers who know Spanish, Russian, Portuguese, Chinese and local languages.

Verifying the results

To verify the results, the data were subjected to statistical tests.

One sample t-tests

```
EF_tTest <- t.test(bi_lang_pct$English_French, mu=23)
Results:
t = 1.8156, df = 5, p-value = 0.1291
alternative hypothesis: true mean is not equal to 23
95 percent confidence interval:
 22.49681 25.92319
sample estimates:
mean of x
    24.21

EA_tTest <- t.test(bi_lang_pct$English_Arabic, mu=14)
Results:
t = -0.72481, df = 5, p-value = 0.5011
alternative hypothesis: true mean is not equal to 14
95 percent confidence interval:
 10.49916 15.96084
sample estimates:
mean of x
    13.23

ES_tTest <- t.test(bi_lang_pct$English_Spanish, mu=7)
Results:
t = 0.52382, df = 5, p-value = 0.6228
alternative hypothesis: true mean is not equal to 7
95 percent confidence interval:
 6.739507 7.393826
sample estimates:
mean of x
 7.066667

EC_tTest <- t.test(bi_lang_pct$English_Chinese, mu=1)
Results:
t = -0.070641, df = 5, p-value = 0.9464
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 0.5637894 1.4128773
sample estimates:
mean of x
0.9883333

ER_tTest <- t.test(bi_lang_pct$English_Russian, mu=2.5)
Results:
t = -0.40725, df = 5, p-value = 0.7007
alternative hypothesis: true mean is not equal to 2.5
95 percent confidence interval:
 1.963787 2.889546
sample estimates:
mean of x
 2.426667

EP_tTest <- t.test(bi_lang_pct$English_Portuguese, mu=2)
Results:
t = 3.9341, df = 5, p-value = 0.01102
alternative hypothesis: true mean is not equal to 2
95 percent confidence interval:
 2.251855 3.201479
sample estimates:
mean of x
 2.726667

EL_tTest <- t.test(bi_lang_pct$English_local, mu=2.5)
Results:
t = 3.6344, df = 5, p-value = 0.01499
```

```
alternative hypothesis: true mean is not equal to 2.5
95 percent confidence interval:
 2.61708 3.18292
sample estimates:
mean of x
  2.9

EFA_tTest <- t.test(bi_lang_pct$English_French_Arabic, mu=33)
Results:
t = 0.77346, df = 5, p-value = 0.4742
alternative hypothesis: true mean is not equal to 33
95 percent confidence interval:
 31.92346 35.00321
sample estimates:
mean of x
 33.46333

EFS_tTest <- t.test(bi_lang_pct$English_French_Spanish, mu=26)
Results:
t = 1.2309, df = 5, p-value = 0.2731
alternative hypothesis: true mean is not equal to 26
95 percent confidence interval:
 25.37604 27.77062
sample estimates:
mean of x
 26.57333

EFAS_tTest <- t.test(bi_lang_pct$English_French_Arabic_Spanish, mu=35)
Results:
t = 0.88356, df = 5, p-value = 0.4174
alternative hypothesis: true mean is not equal to 35
95 percent confidence interval:
 34.02942 36.98725
sample estimates:
mean of x
 35.50833
```

The results reject the assumptions on the expected value for each of the variables, i.e, bilingual and multilingual skills. The assumptions were based on the values from previous table and plot.

<u>A two-sample z-test</u>

The near likeness of the proportion of job ads seeking English-Portuguese speakers and that of those seeking English-local language speakers prompted a z-test of the two samples based on the frequencies by month. This would determine whether indeed there are differences in the demand between the two.

The test involved filtering the existing dataset and then creating a new one for the analysis.

```
bidem_jan16 <- filter(rwjAll_bil16, month == 1)
jan16_EP <- sum(bidem_jan16$EP, na.rm = TRUE)
jan16_EL <- sum(bidem_jan16$EL, na.rm = TRUE)

bidem_feb16 <- filter(rwjAll_bil16, month == 2)
feb16_EP <- sum(bidem_feb16$EP, na.rm = TRUE)
feb16_EL <- sum(bidem_feb16$EL, na.rm = TRUE)

bidem_mar16 <- filter(rwjAll_bil16, month == 3)
mar16_EP <- sum(bidem_mar16$EP, na.rm = TRUE)
mar16_EL <- sum(bidem_mar16$EL, na.rm = TRUE)

bidem_apr16 <- filter(rwjAll_bil16, month == 4)
apr16_EP <- sum(bidem_apr16$EP, na.rm = TRUE)
```

```
apr16_EL <- sum(bidem_apr16$EL, na.rm = TRUE)

bidem_may16 <- filter(rwjAll_bil16, month == 5)
may16_EP <- sum(bidem_may16$EP, na.rm = TRUE)
may16_EL <- sum(bidem_may16$EL, na.rm = TRUE)

bidem_jun16 <- filter(rwjAll_bil16, month == 6)
jun16_EP <- sum(bidem_jun16$EP, na.rm = TRUE)
jun16_EL <- sum(bidem_jun16$EL, na.rm = TRUE)

bidem_jan15 <- filter(rwjAll_bil15, month == 1)
jan15_EP <- sum(bidem_jan15$EP, na.rm = TRUE)
jan15_EL <- sum(bidem_jan15$EL, na.rm = TRUE)

bidem_feb15 <- filter(rwjAll_bil15, month == 2)
feb15_EP <- sum(bidem_feb15$EP, na.rm = TRUE)
feb15_EL <- sum(bidem_feb15$EL, na.rm = TRUE)

bidem_mar15 <- filter(rwjAll_bil15, month == 3)
mar15_EP <- sum(bidem_mar15$EP, na.rm = TRUE)
mar15_EL <- sum(bidem_mar15$EL, na.rm = TRUE)

bidem_apr15 <- filter(rwjAll_bil15, month == 4)
apr15_EP <- sum(bidem_apr15$EP, na.rm = TRUE)
apr15_EL <- sum(bidem_apr15$EL, na.rm = TRUE)

bidem_may15 <- filter(rwjAll_bil15, month == 5)
may15_EP <- sum(bidem_may15$EP, na.rm = TRUE)
may15_EL <- sum(bidem_may15$EL, na.rm = TRUE)

bidem_jun15 <- filter(rwjAll_bil15, month == 6)
jun15_EP <- sum(bidem_jun15$EP, na.rm = TRUE)
jun15_EL <- sum(bidem_jun15$EL, na.rm = TRUE)

bidem_jul15 <- filter(rwjAll_bil15, month == 7)
jul15_EP <- sum(bidem_jul15$EP, na.rm = TRUE)
jul15_EL <- sum(bidem_jul15$EL, na.rm = TRUE)

bidem_aug15 <- filter(rwjAll_bil15, month == 8)
aug15_EP <- sum(bidem_aug15$EP, na.rm = TRUE)
aug15_EL <- sum(bidem_aug15$EL, na.rm = TRUE)

bidem_sep15 <- filter(rwjAll_bil15, month == 9)
sep15_EP <- sum(bidem_sep15$EP, na.rm = TRUE)
sep15_EL <- sum(bidem_sep15$EL, na.rm = TRUE)

bidem_oct15 <- filter(rwjAll_bil15, month == 10)
oct15_EP <- sum(bidem_oct15$EP, na.rm = TRUE)
oct15_EL <- sum(bidem_oct15$EL, na.rm = TRUE)

bidem_nov15 <- filter(rwjAll_bil15, month == 11)
nov15_EP <- sum(bidem_nov15$EP, na.rm = TRUE)
nov15_EL <- sum(bidem_nov15$EL, na.rm = TRUE)

bidem_dec15 <- filter(rwjAll_bil15, month == 12)
dec15_EP <- sum(bidem_dec15$EP, na.rm = TRUE)
dec15_EL <- sum(bidem_dec15$EL, na.rm = TRUE)

bidem_jan14 <- filter(rwjAll_bil14, month == 1)
jan14_EP <- sum(bidem_jan14$EP, na.rm = TRUE)
jan14_EL <- sum(bidem_jan14$EL, na.rm = TRUE)

bidem_feb14 <- filter(rwjAll_bil14, month == 2)
feb14_EP <- sum(bidem_feb14$EP, na.rm = TRUE)
feb14_EL <- sum(bidem_feb14$EL, na.rm = TRUE)

bidem_mar14 <- filter(rwjAll_bil14, month == 3)
mar14_EP <- sum(bidem_mar14$EP, na.rm = TRUE)
```

```
mar14_EL <- sum(bidem_mar14$EL, na.rm = TRUE)

bidem_apr14 <- filter(rwjAll_bil14, month == 4)
apr14_EP <- sum(bidem_apr14$EP, na.rm = TRUE)
apr14_EL <- sum(bidem_apr14$EL, na.rm = TRUE)

bidem_may14 <- filter(rwjAll_bil14, month == 5)
may14_EP <- sum(bidem_may14$EP, na.rm = TRUE)
may14_EL <- sum(bidem_may14$EL, na.rm = TRUE)

bidem_jun14 <- filter(rwjAll_bil14, month == 6)
jun14_EP <- sum(bidem_jun14$EP, na.rm = TRUE)
jun14_EL <- sum(bidem_jun14$EL, na.rm = TRUE)

bidem_jul14 <- filter(rwjAll_bil14, month == 7)
jul14_EP <- sum(bidem_jul14$EP, na.rm = TRUE)
jul14_EL <- sum(bidem_jul14$EL, na.rm = TRUE)

bidem_aug14 <- filter(rwjAll_bil14, month == 8)
aug14_EP <- sum(bidem_aug14$EP, na.rm = TRUE)
aug14_EL <- sum(bidem_aug14$EL, na.rm = TRUE)

bidem_sep14 <- filter(rwjAll_bil14, month == 9)
sep14_EP <- sum(bidem_sep14$EP, na.rm = TRUE)
sep14_EL <- sum(bidem_sep14$EL, na.rm = TRUE)

bidem_oct14 <- filter(rwjAll_bil14, month == 10)
oct14_EP <- sum(bidem_oct14$EP, na.rm = TRUE)
oct14_EL <- sum(bidem_oct14$EL, na.rm = TRUE)

bidem_nov14 <- filter(rwjAll_bil14, month == 11)
nov14_EP <- sum(bidem_nov14$EP, na.rm = TRUE)
nov14_EL <- sum(bidem_nov14$EL, na.rm = TRUE)

bidem_dec14 <- filter(rwjAll_bil14, month == 12)
dec14_EP <- sum(bidem_dec14$EP, na.rm = TRUE)
dec14_EL <- sum(bidem_dec14$EL, na.rm = TRUE)

bidem_jan13 <- filter(rwjAll_bil13, month == 1)
jan13_EP <- sum(bidem_jan13$EP, na.rm = TRUE)
jan13_EL <- sum(bidem_jan13$EL, na.rm = TRUE)

bidem_feb13 <- filter(rwjAll_bil13, month == 2)
feb13_EP <- sum(bidem_feb13$EP, na.rm = TRUE)
feb13_EL <- sum(bidem_feb13$EL, na.rm = TRUE)

bidem_mar13 <- filter(rwjAll_bil13, month == 3)
mar13_EP <- sum(bidem_mar13$EP, na.rm = TRUE)
mar13_EL <- sum(bidem_mar13$EL, na.rm = TRUE)

bidem_apr13 <- filter(rwjAll_bil13, month == 4)
apr13_EP <- sum(bidem_apr13$EP, na.rm = TRUE)
apr13_EL <- sum(bidem_apr13$EL, na.rm = TRUE)

bidem_may13 <- filter(rwjAll_bil13, month == 5)
may13_EP <- sum(bidem_may13$EP, na.rm = TRUE)
may13_EL <- sum(bidem_may13$EL, na.rm = TRUE)

bidem_jun13 <- filter(rwjAll_bil13, month == 6)
jun13_EP <- sum(bidem_jun13$EP, na.rm = TRUE)
jun13_EL <- sum(bidem_jun13$EL, na.rm = TRUE)

bidem_jul13 <- filter(rwjAll_bil13, month == 7)
jul13_EP <- sum(bidem_jul13$EP, na.rm = TRUE)
jul13_EL <- sum(bidem_jul13$EL, na.rm = TRUE)

bidem_aug13 <- filter(rwjAll_bil13, month == 8)
aug13_EP <- sum(bidem_aug13$EP, na.rm = TRUE)
```

```
aug13_EL <- sum(bidem_aug13$EL, na.rm = TRUE)

bidem_sep13 <- filter(rwjAll_bil13, month == 9)
sep13_EP <- sum(bidem_sep13$EP, na.rm = TRUE)
sep13_EL <- sum(bidem_sep13$EL, na.rm = TRUE)

bidem_oct13 <- filter(rwjAll_bil13, month == 10)
oct13_EP <- sum(bidem_oct13$EP, na.rm = TRUE)
oct13_EL <- sum(bidem_oct13$EL, na.rm = TRUE)

bidem_nov13 <- filter(rwjAll_bil13, month == 11)
nov13_EP <- sum(bidem_nov13$EP, na.rm = TRUE)
nov13_EL <- sum(bidem_nov13$EL, na.rm = TRUE)

bidem_dec13 <- filter(rwjAll_bil13, month == 12)
dec13_EP <- sum(bidem_dec13$EP, na.rm = TRUE)
dec13_EL <- sum(bidem_dec13$EL, na.rm = TRUE)

bidem_jan12 <- filter(rwjAll_bil12, month == 1)
jan12_EP <- sum(bidem_jan12$EP, na.rm = TRUE)
jan12_EL <- sum(bidem_jan12$EL, na.rm = TRUE)

bidem_feb12 <- filter(rwjAll_bil12, month == 2)
feb12_EP <- sum(bidem_feb12$EP, na.rm = TRUE)
feb12_EL <- sum(bidem_feb12$EL, na.rm = TRUE)

bidem_mar12 <- filter(rwjAll_bil12, month == 3)
mar12_EP <- sum(bidem_mar12$EP, na.rm = TRUE)
mar12_EL <- sum(bidem_mar12$EL, na.rm = TRUE)

bidem_apr12 <- filter(rwjAll_bil12, month == 4)
apr12_EP <- sum(bidem_apr12$EP, na.rm = TRUE)
apr12_EL <- sum(bidem_apr12$EL, na.rm = TRUE)

bidem_may12 <- filter(rwjAll_bil12, month == 5)
may12_EP <- sum(bidem_may12$EP, na.rm = TRUE)
may12_EL <- sum(bidem_may12$EL, na.rm = TRUE)

bidem_jun12 <- filter(rwjAll_bil12, month == 6)
jun12_EP <- sum(bidem_jun12$EP, na.rm = TRUE)
jun12_EL <- sum(bidem_jun12$EL, na.rm = TRUE)

bidem_jul12 <- filter(rwjAll_bil12, month == 7)
jul12_EP <- sum(bidem_jul12$EP, na.rm = TRUE)
jul12_EL <- sum(bidem_jul12$EL, na.rm = TRUE)

bidem_aug12 <- filter(rwjAll_bil12, month == 8)
aug12_EP <- sum(bidem_aug12$EP, na.rm = TRUE)
aug12_EL <- sum(bidem_aug12$EL, na.rm = TRUE)

bidem_sep12 <- filter(rwjAll_bil12, month == 9)
sep12_EP <- sum(bidem_sep12$EP, na.rm = TRUE)
sep12_EL <- sum(bidem_sep12$EL, na.rm = TRUE)

bidem_oct12 <- filter(rwjAll_bil12, month == 10)
oct12_EP <- sum(bidem_oct12$EP, na.rm = TRUE)
oct12_EL <- sum(bidem_oct12$EL, na.rm = TRUE)

bidem_nov12 <- filter(rwjAll_bil12, month == 11)
nov12_EP <- sum(bidem_nov12$EP, na.rm = TRUE)
nov12_EL <- sum(bidem_nov12$EL, na.rm = TRUE)

bidem_dec12 <- filter(rwjAll_bil12, month == 12)
dec12_EP <- sum(bidem_dec12$EP, na.rm = TRUE)
dec12_EL <- sum(bidem_dec12$EL, na.rm = TRUE)

bidem_mar11 <- filter(rwjAll_bil11, month == 3)
mar11_EP <- sum(bidem_mar11$EP, na.rm = TRUE)
```

```
mar11_EL <- sum(bidem_mar11$EL, na.rm = TRUE)

bidem_apr11 <- filter(rwjAll_bil11, month == 4)
apr11_EP <- sum(bidem_apr11$EP, na.rm = TRUE)
apr11_EL <- sum(bidem_apr11$EL, na.rm = TRUE)

bidem_may11 <- filter(rwjAll_bil11, month == 5)
may11_EP <- sum(bidem_may11$EP, na.rm = TRUE)
may11_EL <- sum(bidem_may11$EL, na.rm = TRUE)

bidem_jun11 <- filter(rwjAll_bil11, month == 6)
jun11_EP <- sum(bidem_jun11$EP, na.rm = TRUE)
jun11_EL <- sum(bidem_jun11$EL, na.rm = TRUE)

bidem_jul11 <- filter(rwjAll_bil11, month == 7)
jul11_EP <- sum(bidem_jul11$EP, na.rm = TRUE)
jul11_EL <- sum(bidem_jul11$EL, na.rm = TRUE)

bidem_aug11 <- filter(rwjAll_bil11, month == 8)
aug11_EP <- sum(bidem_aug11$EP, na.rm = TRUE)
aug11_EL <- sum(bidem_aug11$EL, na.rm = TRUE)

bidem_sep11 <- filter(rwjAll_bil11, month == 9)
sep11_EP <- sum(bidem_sep11$EP, na.rm = TRUE)
sep11_EL <- sum(bidem_sep11$EL, na.rm = TRUE)

bidem_oct11 <- filter(rwjAll_bil11, month == 10)
oct11_EP <- sum(bidem_oct11$EP, na.rm = TRUE)
oct11_EL <- sum(bidem_oct11$EL, na.rm = TRUE)

bidem_nov11 <- filter(rwjAll_bil11, month == 11)
nov11_EP <- sum(bidem_nov11$EP, na.rm = TRUE)
nov11_EL <- sum(bidem_nov11$EL, na.rm = TRUE)

bidem_dec11 <- filter(rwjAll_bil11, month == 12)
dec11_EP <- sum(bidem_dec11$EP, na.rm = TRUE)
dec11_EL <- sum(bidem_dec11$EL, na.rm = TRUE)

Year <- c(2011, 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2012, 2012, 2012,
2012, 2012, 2012, 2012, 2012, 2012, 2012, 2012, 2012, 2013, 2013, 2013, 2013, 2013, 2013,
2013, 2013, 2013, 2013, 2013, 2013, 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014,
2014, 2014, 2014, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015,
2016, 2016, 2016, 2016, 2016, 2016)

Month <- c(03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 01, 02, 03, 04, 05, 06, 07, 08, 09,
10, 11, 12, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 01, 02, 03, 04, 05, 06, 07,
08, 09, 10, 11, 12, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 01, 02, 03, 04, 05,
06)

English_Portuguese <- c(mar11_EP, apr11_EP, may11_EP, jun11_EP, jul11_EP, aug11_EP,
sep11_EP, oct11_EP, nov11_EP, dec11_EP, jan12_EP, feb12_EP, mar12_EP, apr12_EP, may12_EP,
jun12_EP, jul12_EP, aug12_EP, sep12_EP, oct12_EP, nov12_EP, dec12_EP, jan13_EP, feb13_EP,
mar13_EP, apr13_EP, may13_EP, jun13_EP, jul13_EP, aug13_EP, sep13_EP, oct13_EP, nov13_EP,
dec13_EP, jan14_EP, feb14_EP, mar14_EP, apr14_EP, may14_EP, jun14_EP, jul14_EP, aug14_EP,
sep14_EP, oct14_EP, nov14_EP, dec14_EP, jan15_EP, feb15_EP, mar15_EP, apr15_EP, may15_EP,
jun15_EP, jul15_EP, aug15_EP, sep15_EP, oct15_EP, nov15_EP, dec15_EP, jan16_EP, feb16_EP,
mar16_EP, apr16_EP, may16_EP, jun16_EP)

English_local <- c(mar11_EL, apr11_EL, may11_EL, jun11_EL, jul11_EL, aug11_EL, sep11_EL,
oct11_EL, nov11_EL, dec11_EL, jan12_EL, feb12_EL, mar12_EL, apr12_EL, may12_EL, jun12_EL,
jul12_EL, aug12_EL, sep12_EL, oct12_EL, nov12_EL, dec12_EL, jan13_EL, feb13_EL, mar13_EL,
apr13_EL, may13_EL, jun13_EL, jul13_EL, aug13_EL, sep13_EL, oct13_EL, nov13_EL, dec13_EL,
jan14_EL, feb14_EL, mar14_EL, apr14_EL, may14_EL, jun14_EL, jul14_EL, aug14_EL, sep14_EL,
oct14_EL, nov14_EL, dec14_EL, jan15_EL, feb15_EL, mar15_EL, apr15_EL, may15_EL, jun15_EL,
jul15_EL, aug15_EL, sep15_EL, oct15_EL, nov15_EL, dec15_EL, jan16_EL, feb16_EL, mar16_EL,
apr16_EL, may16_EL, jun16_EL)

zTestdf <- data.frame(Year, Month, English_Portuguese, English_local)
```

The below shows the data to be used for the z-test, which involved calling the BSDA package.

Year	Month	English_ Portuguese	English_ Local
2011	3	0	0
2011	4	2	0
2011	5	16	8
2011	6	44	42
2011	7	60	68
2011	8	112	114
2011	9	60	56
2011	10	102	48
2011	11	80	70
2011	12	98	52
2012	1	74	48
2012	2	144	76
2012	3	74	84
2012	4	72	90
2012	5	32	44
2012	6	130	120
2012	7	56	90
2012	8	62	88
2012	9	92	136
2012	10	80	78
2012	11	82	94
2012	12	56	82
2013	1	106	82
2013	2	80	74
2013	3	90	110
2013	4	96	120
2013	5	110	112
2013	6	126	132
2013	7	112	130
2013	8	100	86
2013	9	146	118

2013	10	84	70
2013	11	46	82
2013	12	90	102
2014	1	128	120
2014	2	70	92
2014	3	72	90
2014	4	118	90
2014	5	78	134
2014	6	78	88
2014	7	102	78
2014	8	88	66
2014	9	114	80
2014	10	98	154
2014	11	110	78
2014	12	92	118
2015	1	68	80
2015	2	70	110
2015	3	78	132
2015	4	80	72
2015	5	64	94
2015	6	0	0
2015	7	140	196
2015	8	90	100
2015	9	60	72
2015	10	98	142
2015	11	48	112
2015	12	96	126
2016	1	78	94
2016	2	106	90
2016	3	88	128
2016	4	76	106
2016	5	80	118
2016	6	58	62

```
library(BSDA)

z.test(zTestdf$English_Portuguese, zTestdf$English_local, sigma.x =
sd(zTestdf$English_Portuguese), sigma.y = sd(zTestdf$English_local), conf.level = 0.95)
Results:
z = -1.2629, p-value = 0.2066
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -19.458401   4.208401
sample estimates:
mean of x mean of y
   81.875    89.500
```

The two variables were also tested for correlation:

```
cor(zTestdf$English_Portuguese, zTestdf$English_local)
[1] 0.6733323
```

The results suggest that there is a difference between the demand for English-Portuguese speakers and that for English-local language speakers, and that there is a relative association between the two.

Linear regression

The dataset showing annual trends in the demand for English speakers with foreign language skills (see the first table in the "Analyzing the data" section) has prompted questions whether certain variables could predict these trends.

The assumption is that foreign aid, in terms of commitments or disbursements or both, may predict those patterns. To verify such an assumption, data on foreign aid, technically known as official development assistance, were extracted from the website of the Organization for Economic Cooperation and Development.

OECD tracks foreign aid committed and disbursed by donor countries. It has data on ODA through 2014, and allows downloading of such data from its statistics page, http://stats.oecd.org and saving them into a csv file.

To prepare the data for linear modeling, the extracted datasets on foreign aid commitments and disbursements were filtered according to the below criteria:

- Year, from 2011 to 2014.
- French-speaking aid-recipient countries.
- Arabic-speaking aid-recipient countries.
- Values on "total commitments" for the dataset on ODA commitments.
- Values on "grants, total", "technical cooperation" and "humanitarian aid" for the dataset on ODA disbursements.

```
French_speaking_countries <- c("Benin", "Burkina Faso", "Burundi", "Cameroon", "Central
African Republic", "Chad", "Comoros", "Democratic Republic of the Congo", "Congo", "Côte
d'Ivoire", "Djibouti", "Equatorial Guinea", "Gabon", "Guinea", "Madagascar", "Mali",
"Mauritius", "Morocco", "Niger", "Rwanda", "Senegal", "Togo", "Haiti", "Vanuatu",
"Seychelles")
```

```
Arabic_speaking_countries <- c("Benin", "Chad", "Comoros", "Djibouti", "Egypt",
"Eritrea", "Libya", "Mauritania", "Morocco", "Sudan", "Tunisia", "Tanzania", "Iraq",
"Jordan", "Lebanon", "Oman", "West Bank and Gaza Strip", "Syrian Arab Republic", "Yemen")

ODA_commitments_11_14 <- filter(ODA_commitments, Year >= 2011)


ODA_commitments_French_speaking <- ODA_commitments_11_14[ODA_commitments_11_14$Recipient
%in% French_speaking_countries,]
ODA_commitments_French_speaking <- filter(ODA_commitments_French_speaking, Aid.type ==
"Total Commitments")
ODA_commitments_French_speaking <- filter(ODA_commitments_French_speaking, Donor == "All
Donors, Total")

ODA_commitments_Arabic_speaking <- ODA_commitments_11_14[ODA_commitments_11_14$Recipient
%in% Arabic_speaking_countries,]
ODA_commitments_Arabic_speaking <- filter(ODA_commitments_Arabic_speaking, Aid.type ==
"Total Commitments")
ODA_commitments_Arabic_speaking <- filter(ODA_commitments_Arabic_speaking, Donor == "All
Donors, Total")

ODA_disbursements_11_14 <- filter(ODA_disbursements, Year >= 2011)

ODA_disbursements_French_speaking <-
ODA_disbursements_11_14[ODA_disbursements_11_14$Recipient %in%
French_speaking_countries,]
ODA_disbursements_French_speaking <- filter(ODA_disbursements_French_speaking, Donor ==
"All Donors, Total")
ODA_disbursements_type1 <- filter(ODA_disbursements_French_speaking, Aid.type == "Grants,
Total")
ODA_disbursements_type2 <- filter(ODA_disbursements_French_speaking, Aid.type ==
"Technical Cooperation")
ODA_disbursements_type3 <- filter(ODA_disbursements_French_speaking, Aid.type ==
"Humanitarian Aid")

ODA_disbursements_Arabic_speaking <-
ODA_disbursements_11_14[ODA_disbursements_11_14$Recipient %in%
Arabic_speaking_countries,]
ODA_disbursements_Arabic_speaking <- filter(ODA_disbursements_Arabic_speaking, Donor ==
"All Donors, Total")
ODA_disbursements_type1 <- filter(ODA_disbursements_Arabic_speaking, Aid.type == "Grants,
Total")
ODA_disbursements_type2 <- filter(ODA_disbursements_Arabic_speaking, Aid.type ==
"Technical Cooperation")
ODA_disbursements_type3 <- filter(ODA_disbursements_Arabic_speaking, Aid.type ==
"Humanitarian Aid")
ODA_disbursements_Arabic_speaking <- rbind(ODA_disbursements_type1,
ODA_disbursements_type2, ODA_disbursements_type3)
```

Unnecessary columns were removed.

```
ODA_commitments_French_speaking$DONOR <- NULL
ODA_commitments_French_speaking$RECIPIENT <- NULL
ODA_commitments_French_speaking$PART <- NULL
ODA_commitments_French_speaking$AIDTYPE <- NULL
ODA_commitments_French_speaking$DATATYPE <- NULL
ODA_commitments_French_speaking$TIME <- NULL
ODA_commitments_French_speaking$Unit.Code <- NULL
ODA_commitments_French_speaking$PowerCode.Code <- NULL
ODA_commitments_French_speaking$Reference.Period <- NULL
ODA_commitments_French_speaking$Reference.Period.Code <- NULL
ODA_commitments_French_speaking$Flag.Codes <- NULL
ODA_commitments_French_speaking$Flags <- NULL
ODA_commitments_French_speaking$Part <- NULL

ODA_commitments_Arabic_speaking$DONOR <- NULL
ODA_commitments_Arabic_speaking$RECIPIENT <- NULL
ODA_commitments_Arabic_speaking$PART <- NULL
```

```
ODA_commitments_Arabic_speaking$AIDTYPE <- NULL
ODA_commitments_Arabic_speaking$DATATYPE <- NULL
ODA_commitments_Arabic_speaking$TIME <- NULL
ODA_commitments_Arabic_speaking$Unit.Code <- NULL
ODA_commitments_Arabic_speaking$PowerCode.Code <- NULL
ODA_commitments_Arabic_speaking$Reference.Period <- NULL
ODA_commitments_Arabic_speaking$Reference.Period.Code <- NULL
ODA_commitments_Arabic_speaking$Flag.Codes <- NULL
ODA_commitments_Arabic_speaking$Flags <- NULL
ODA_commitments_Arabic_speaking$Part <- NULL

ODA_disbursements_French_speaking <- rbind(ODA_disbursements_type1,
ODA_disbursements_type2, ODA_disbursements_type3)
ODA_disbursements_French_speaking$DONOR <- NULL
ODA_disbursements_French_speaking$RECIPIENT <- NULL
ODA_disbursements_French_speaking$PART <- NULL
ODA_disbursements_French_speaking$AIDTYPE <- NULL
ODA_disbursements_French_speaking$DATATYPE <- NULL
ODA_disbursements_French_speaking$TIME <- NULL
ODA_disbursements_French_speaking$Unit.Code <- NULL
ODA_disbursements_French_speaking$PowerCode.Code <- NULL
ODA_disbursements_French_speaking$Reference.Period <- NULL
ODA_disbursements_French_speaking$Reference.Period.Code <- NULL
ODA_disbursements_French_speaking$Flag.Codes <- NULL
ODA_disbursements_French_speaking$Flags <- NULL
ODA_disbursements_French_speaking$Part <- NULL

ODA_disbursements_Arabic_speaking$DONOR <- NULL
ODA_disbursements_Arabic_speaking$RECIPIENT <- NULL
ODA_disbursements_Arabic_speaking$PART <- NULL
ODA_disbursements_Arabic_speaking$AIDTYPE <- NULL
ODA_disbursements_Arabic_speaking$DATATYPE <- NULL
ODA_disbursements_Arabic_speaking$TIME <- NULL
ODA_disbursements_Arabic_speaking$Unit.Code <- NULL
ODA_disbursements_Arabic_speaking$PowerCode.Code <- NULL
ODA_disbursements_Arabic_speaking$Reference.Period <- NULL
ODA_disbursements_Arabic_speaking$Reference.Period.Code <- NULL
ODA_disbursements_Arabic_speaking$Flag.Codes <- NULL
ODA_disbursements_Arabic_speaking$Flags <- NULL
ODA_disbursements_Arabic_speaking$Part <- NULL
```

The filtered datasets were split into separate datasets to determine the annual trends.

```
ODA_commitments_French_speaking_11 <- filter(ODA_commitments_French_speaking, Year ==
2011)
ODA_commitments_French_speaking_12 <- filter(ODA_commitments_French_speaking, Year ==
2012)
ODA_commitments_French_speaking_13 <- filter(ODA_commitments_French_speaking, Year ==
2013)
ODA_commitments_French_speaking_14 <- filter(ODA_commitments_French_speaking, Year ==
2014)

ODA_commitments_Arabic_speaking_11 <- filter(ODA_commitments_Arabic_speaking, Year ==
2011)
ODA_commitments_Arabic_speaking_12 <- filter(ODA_commitments_Arabic_speaking, Year ==
2012)
ODA_commitments_Arabic_speaking_13 <- filter(ODA_commitments_Arabic_speaking, Year ==
2013)
ODA_commitments_Arabic_speaking_14 <- filter(ODA_commitments_Arabic_speaking, Year ==
2014)

ODA_disbursements_French_speaking_11 <- filter(ODA_disbursements_French_speaking, Year ==
2011)
ODA_disbursements_French_speaking_12 <- filter(ODA_disbursements_French_speaking, Year ==
2012)
```

```
ODA_disbursements_French_speaking_13 <- filter(ODA_disbursements_French_speaking, Year ==
2013)
ODA_disbursements_French_speaking_14 <- filter(ODA_disbursements_French_speaking, Year ==
2014)

ODA_disbursements_Arabic_speaking_11 <- filter(ODA_disbursements_Arabic_speaking, Year ==
2011)
ODA_disbursements_Arabic_speaking_12 <- filter(ODA_disbursements_Arabic_speaking, Year ==
2012)
ODA_disbursements_Arabic_speaking_13 <- filter(ODA_disbursements_Arabic_speaking, Year ==
2013)
ODA_disbursements_Arabic_speaking_14 <- filter(ODA_disbursements_Arabic_speaking, Year ==
2014)
```

The summed ODA amounts populated the final dataset that was used for the linear regression.

```
ESC_FS_11 <- sum(ODA_commitments_French_speaking_11$Value)
ESC_FS_12 <- sum(ODA_commitments_French_speaking_12$Value)
ESC_FS_13 <- sum(ODA_commitments_French_speaking_13$Value)
ESC_FS_14 <- sum(ODA_commitments_French_speaking_14$Value)

ESC_AS_11 <- sum(ODA_commitments_Arabic_speaking_11$Value)
ESC_AS_12 <- sum(ODA_commitments_Arabic_speaking_12$Value)
ESC_AS_13 <- sum(ODA_commitments_Arabic_speaking_13$Value)
ESC_AS_14 <- sum(ODA_commitments_Arabic_speaking_14$Value)

ESD_FS_11 <- sum(ODA_disbursements_French_speaking_11$Value)
ESD_FS_12 <- sum(ODA_disbursements_French_speaking_12$Value)
ESD_FS_13 <- sum(ODA_disbursements_French_speaking_13$Value)
ESD_FS_14 <- sum(ODA_disbursements_French_speaking_14$Value)

ESD_AS_11 <- sum(ODA_disbursements_Arabic_speaking_11$Value)
ESD_AS_12 <- sum(ODA_disbursements_Arabic_speaking_12$Value)
ESD_AS_13 <- sum(ODA_disbursements_Arabic_speaking_13$Value)
ESD_AS_14 <- sum(ODA_disbursements_Arabic_speaking_14$Value)

Year <- c(2011, 2012, 2013, 2014)

ODAcommitments_FS <- c(ESC_FS_11, ESC_FS_12, ESC_FS_13, ESC_FS_14)
ODAdisbursements_FS <- c(ESD_FS_11, ESD_FS_12, ESD_FS_13, ESD_FS_14)
EnglishFrench_pct <- c(EF_percent_11, EF_percent_12, EF_percent_13, EF_percent_14)
ODAvsEFjobs <- data.frame(Year, ODAcommitments_FS, ODAdisbursements_FS,
EnglishFrench_pct)
ODAvsEFjobs$EnglishFrench_pct <- ODAvsEFjobs$EnglishFrench_pct * 100
ODAvsEFjobs$EnglishFrench_pct <- round(ODAvsEFjobs$EnglishFrench_pct, digits = 2)

ODAcommitments_AS <- c(ESC_AS_11, ESC_AS_12, ESC_AS_13, ESC_AS_14)
ODAdisbursements_AS <- c(ESD_AS_11, ESD_AS_12, ESD_AS_13, ESD_AS_14)
EnglishArabic_pct <- c(EA_percent_11, EA_percent_12, EA_percent_13, EA_percent_14)
ODAvsEAjobs <- data.frame(Year, ODAcommitments_AS, ODAdisbursements_AS,
EnglishArabic_pct)
ODAvsEAjobs <- as.data.frame(ODAvsEAjobs)
ODAvsEAjobs$EnglishArabic_pct <- ODAvsEAjobs$EnglishArabic_pct * 100
ODAvsEAjobs$EnglishArabic_pct <- round(ODAvsEAjobs$EnglishArabic_pct, digits = 2)
```

The following table illustrates the resulting dataset:

Year	ODAcommitments_FS (in US$ millions)	ODAdisbursements_FS (in US$ millions)	EnglishFrench_pct
2011	42640.47	46200.75	26.85
2012	44283.17	44513.33	24.53

2013	41983.15	38207.44	23.86
2014	38298.24	35761.01	21.76

Year	ODAcommitments_ AS (in US$ millions)	ODAdisbursements_ AS (in US$ millions)	EnglishArabic_ Pct
2011	36238.31	36683.82	10.12
2012	47168.99	37559.22	10.56
2013	59993.11	53625.05	13.46
2014	48858.18	57435.2	13.26

Several models looked at whether ODA commitments and disbursements can affect the demand for English-French and English-Arabic speakers in the aid industry.

```
model1 <- lm(EnglishFrench_pct ~ ODAdisbursements_FS + ODAcommitments_FS, data =
ODAvsEFjobs)

summary(model1)

Residuals:
      1       2       3       4
 0.5962 -0.9428  0.8287 -0.4821

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          9.825e+00  1.608e+01   0.611    0.651
ODAdisbursements_FS  4.078e-04  2.965e-04   1.376    0.400
ODAcommitments_FS   -5.656e-05  5.845e-04  -0.097    0.939

Residual standard error: 1.471 on 1 degrees of freedom
Multiple R-squared:  0.836,	Adjusted R-squared:  0.508
F-statistic: 2.549 on 2 and 1 DF,  p-value: 0.405

model2 <- lm(EnglishFrench_pct ~ ODAcommitments_FS, data = ODAvsEFjobs)

summary(model2)

Residuals:
      1       2       3       4
 2.0953 -1.2127 -0.4994 -0.3832

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.8905793 16.9106452  -0.053    0.963
ODAcommitments_FS  0.0006014  0.0004040   1.489    0.275

Residual standard error: 1.769 on 2 degrees of freedom
Multiple R-squared:  0.5256,	Adjusted R-squared:  0.2885
F-statistic: 2.216 on 1 and 2 DF,  p-value: 0.275

model3 <- lm(EnglishFrench_pct ~ ODAdisbursements_FS, data = ODAvsEFjobs)

summary(model3)

Residuals:
      1       2       3       4
 0.6668 -1.0047  0.7488 -0.4110
```

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         8.4271646  5.0107034   1.682   0.2346
ODAdisbursements_FS 0.0003843  0.0001210   3.175   0.0865 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.045 on 2 degrees of freedom
Multiple R-squared:  0.8345,      Adjusted R-squared:  0.7517
F-statistic: 10.08 on 1 and 2 DF,  p-value: 0.08651

model4 <- lm(EnglishArabic_pct ~ ODAdisbursements_AS + ODAcommitments_AS, data =
ODAvsEAjobs)

summary(model4)

Residuals:
       1        2        3        4
 0.09150 -0.11215  0.08669 -0.06604

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.523e+00  5.403e-01   6.521   0.0969 .
ODAdisbursements_AS 1.294e-04  1.353e-05   9.563   0.0663 .
ODAcommitments_AS   4.852e-05  1.496e-05   3.244   0.1904
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1812 on 1 degrees of freedom
Multiple R-squared:  0.9964,      Adjusted R-squared:  0.9893
F-statistic: 140.2 on 2 and 1 DF,  p-value: 0.05962

model5 <- lm(EnglishArabic_pct ~ ODAcommitments_AS, data = ODAvsEAjobs)

summary(model5)

Residuals:
       1        2        3        4
 0.01898 -1.15754 -0.15409  1.29265

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       4.742e+00  3.570e+00   1.328    0.315
ODAcommitments_AS 1.479e-04  7.315e-05   2.022    0.181

Residual standard error: 1.232 on 2 degrees of freedom
Multiple R-squared:  0.6714,      Adjusted R-squared:  0.5072
F-statistic: 4.087 on 1 and 2 DF,  p-value: 0.1806

model6 <- lm(EnglishArabic_pct ~ ODAdisbursements_AS, data = ODAvsEAjobs)

summary(model6)

Residuals:
      1       2       3       4
-0.1883  0.1117  0.4429 -0.3663

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.443e+00  1.104e+00   4.025   0.0566 .
ODAdisbursements_AS 1.599e-04  2.336e-05   6.844   0.0207 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4349 on 2 degrees of freedom
Multiple R-squared:  0.959,       Adjusted R-squared:  0.9386
F-statistic: 46.84 on 1 and 2 DF,  p-value: 0.02069
```

Correlation among variables were also carried out.

```
cor(ODAvsEFjobs$EnglishFrench_pct, ODAvsEFjobs$ODAdisbursements_FS)
[1] 0.9134854
cor(ODAvsEFjobs$EnglishFrench_pct, ODAvsEFjobs$ODAcommitments_FS)
[1] 0.725014

cor(ODAvsEAjobs$EnglishArabic_pct, ODAvsEAjobs$ODAdisbursements_AS)
[1] 0.9793095
cor(ODAvsEAjobs$EnglishArabic_pct, ODAvsEAjobs$ODAcommitments_AS)
[1] 0.8194168
```

The results suggest that there is a significant relationship and a high correlation between foreign aid disbursements and the demand for English-Arabic job candidates. Correlation between foreign aid commitment and the demand for English-Arabic job candidates as well as between foreign aid (commitments or disbursements) and the demand for English-French speakers is high but the relationship is not significant. These however are not conclusive given the small sample.

Takeaways

1. If we are to rank the demand for bilingual English speakers in global development and humanitarian relief, those who know French would top the list, followed by English-Arabic speakers. Here's a quick look, together with the extent of the demand based on the 2016 average.

English speakers who know…	Demand
French	~ 24 in 100 jobs
Arabic	~ 17 in 100 jobs
Spanish	~ 8 in 100 jobs
Russian	~ 3 in 100 jobs
Local languages	~ 3 in 100 jobs
Portuguese	~ 2 in 100 jobs
Chinese	~ 2 in 100 jobs

2. English speakers who know French can see their job market access increase significantly if they are also proficient in Arabic (about 11 percentage points in 2016). Meanwhile, English-French speakers who know Spanish can only see a slight increase (about 2 percentage points in 2016).

3. The demand for English-Arabic speakers is slowly catching up with that for English-French speakers, as the last six years have seen a steady increase of the former while the latter has suffered a decline.

4. Foreign aid disbursements appear to affect trends in the demand for English-Arab speakers in the aid industry. The same cannot be said for English-French speakers. The pattern though is inconclusive given the small sample for this study.

Recommendations

1. As a job seeker in global development and humanitarian relief, if there's one foreign language that you plan to acquire, invest your time in being proficient in French.

2. If you're already fluent in French and want to know another foreign language, consider learning Arabic, as 35 in 100 jobs seek qualified English-speaking candidates who are have knowledge of French or Arabic, or both these languages.

3. Universities offering courses focused on global development and humanitarian relief should consider concentrating their language training on French and Arabic as this may improve their students' chances of landing jobs in the future. For those without language training components in their academic programs, this study provides a good case for starting such a program or for partnering with language learning centers.