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, those indicating that the position is located in English-speaking donor countries and those that cite having bilingual or multilingual language skills as necessary or preferable qualities of candidates.

Donor countries refer to high-income nations such as the United States, the United Kingdom, Canada, Australia and Ireland which have traditionally provided foreign aid. Apart from donations from private individuals and foundations, foreign aid funds global development and humanitarian relief positions, and these donor countries would likely host organizations that hire people with foreign language skills, whether at headquarters or for their field projects in low-income countries.

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:

- "English" which denotes that the ad asks for applicants to be fluent in or knowledge of English.
- "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.

Binary columns were also created based on keyword searches of the "country" column. The keywords include the "United Kingdom", "Canada", "Australia" and "Ireland". These and the "English" binary columns were combined into one using the for loop function with the ifelse statement.

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

```
rwjall <- as.data.frame(rwjAll)

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

rwjAll$English2 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English2 <- str_match(rwjAll$country, "United States")
rwjAll$English2 <- ifelse(rwjAll$English2 == "United States", 1, 0)
rwjAll$English2 <- unmatrix(rwjAll$English2)
for (i in 1:length(rwjAll$English1))
if (!is.na(rwjAll$English1[i])) rwjAll$English2[i]=NA

rwjAll$English3 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English3 <- str_match(rwjAll$country, "United Kingdom")
```

```
rwjAll$English3 <- ifelse(rwjAll$English3 == "United Kingdom", 1, 0)
rwjAll$English3 <- unmatrix(rwjAll$English3)
for (i in 1:length(rwjAll$English1))
if (!is.na(rwjAll$English1[i])) rwjAll$English3[i]=NA

rwjAll$English4 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English4 <- str_match(rwjAll$country, "Ireland")
rwjAll$English4 <- ifelse(rwjAll$English4 == "Ireland", 1, 0)
rwjAll$English4 <- unmatrix(rwjAll$English4)
for (i in 1:length(rwjAll$English1))
if (!is.na(rwjAll$English1[i])) rwjAll$English4[i]=NA
for (i in 1:length(rwjAll$English3))
if (!is.na(rwjAll$English3[i])) rwjAll$English4[i]=NA

rwjAll$English5 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English5 <- str_match(rwjAll$country, "Canada")
rwjAll$English5 <- ifelse(rwjAll$English5 == "Canada", 1, 0)
rwjAll$English5 <- unmatrix(rwjAll$English5)
for (i in 1:length(rwjAll$English1))
if (!is.na(rwjAll$English1[i])) rwjAll$English5[i]=NA

rwjAll$English6 <- as.vector(c(NA * nrow(rwjAll)))
rwjAll$English6 <- str_match(rwjAll$country, "Australia")
rwjAll$English6 <- ifelse(rwjAll$English6 == "Australia", 1, 0)
rwjAll$English6 <- unmatrix(rwjAll$English6)
for (i in 1:length(rwjAll$English1))
if (!is.na(rwjAll$English1[i])) rwjAll$English6[i]=NA

rwjAll$English_total <- rowSums(rwjAll[, 13:18], na.rm = TRUE)

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$bilingual_all <- rowSums(rwjAll[, 19:26], na.rm = TRUE)
rwjAll$bilingual_all[rwjAll$bilingual_all == 2] <- 1
rwjAll$bilingual_all[rwjAll$bilingual_all == 3] <- 1
```

```
rwjAll$bilingual_all[rwjAll$bilingual_all == 4] <- 1
rwjAll$bilingual_all[rwjAll$bilingual_all == 5] <- 1
rwjAll$bilingual_all[rwjAll$bilingual_all == 6] <- 1
rwjAll$bilingual_all[rwjAll$bilingual_all == 7] <- 1
rwjAll$bilingual_all[rwjAll$bilingual_all == 8] <- 1
rwjAll$French_Arabic <- rowSums(rwjAll[, 20:21], na.rm = TRUE)
rwjAll$French_Spanish <- rowSums(rwjAll[, c("French", "Spanish")], na.rm = TRUE)
rwjAll$French_Arabic_Spanish <- rowSums(rwjAll[, 20:22], na.rm = TRUE)
rwjAll$French_Arabic[rwjAll$French_Arabic==2] <- 1
rwjAll$French_Spanish[rwjAll$French_Spanish==2] <- 1
rwjAll$French_Arabic_Spanish[rwjAll$French_Arabic_Spanish==2] <- 1
rwjAll$French_Arabic_Spanish[rwjAll$French_Arabic_Spanish==3] <- 1

rwjAll_bil <- filter(rwjAll, bilingual_all == 1)
rwjAll_bil <- rwjAll_bil[!duplicated(rwjAll_bil), ]
```

Using the tidyr and dplyr packages, the dataset was filtered (with 74,137 entries after initial filtering) was then filtered by year to compare annual trends.

```
library(tidyr)
library(dplyr)

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.

```
total_EF_16 <- sum(rwjAll_bil16$French, na.rm = TRUE)
EF_percent_16 <- total_EF_16 / nrow(rwjAll_bil16)
total_EA_16 <- sum(rwjAll_bil16$Arabic, na.rm = TRUE)
EA_percent_16 <- total_EA_16 / nrow(rwjAll_bil16)
total_ES_16 <- sum(rwjAll_bil16$Spanish, na.rm = TRUE)
ES_percent_16 <- total_ES_16 / nrow(rwjAll_bil16)
total_ER_16 <- sum(rwjAll_bil16$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)

total_EF_15 <- sum(rwjAll_bil15$French, na.rm = TRUE)
EF_percent_15 <- total_EF_15 / nrow(rwjAll_bil15)
total_EA_15 <- sum(rwjAll_bil15$Arabic, na.rm = TRUE)
EA_percent_15 <- total_EA_15 / nrow(rwjAll_bil15)
total_ES_15 <- sum(rwjAll_bil15$Spanish, na.rm = TRUE)
ES_percent_15 <- total_ES_15 / nrow(rwjAll_bil15)
total_ER_15 <- sum(rwjAll_bil15$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)

total_EF_14 <- sum(rwjAll_bil14$French, na.rm = TRUE)
EF_percent_14 <- total_EF_14 / nrow(rwjAll_bil14)
total_EA_14 <- sum(rwjAll_bil14$Arabic, na.rm = TRUE)
EA_percent_14 <- total_EA_14 / nrow(rwjAll_bil14)
total_ES_14 <- sum(rwjAll_bil14$Spanish, na.rm = TRUE)
ES_percent_14 <- total_ES_14 / nrow(rwjAll_bil14)
total_ER_14 <- sum(rwjAll_bil14$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)

total_EF_13 <- sum(rwjAll_bil13$French, na.rm = TRUE)
EF_percent_13 <- total_EF_13 / nrow(rwjAll_bil13)
total_EA_13 <- sum(rwjAll_bil13$Arabic, na.rm = TRUE)
EA_percent_13 <- total_EA_13 / nrow(rwjAll_bil13)
total_ES_13 <- sum(rwjAll_bil13$Spanish, na.rm = TRUE)
ES_percent_13 <- total_ES_13 / nrow(rwjAll_bil13)
total_ER_13 <- sum(rwjAll_bil13$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)

total_EF_12 <- sum(rwjAll_bil12$French, na.rm = TRUE)
EF_percent_12 <- total_EF_12 / nrow(rwjAll_bil12)
total_EA_12 <- sum(rwjAll_bil12$Arabic, na.rm = TRUE)
EA_percent_12 <- total_EA_12 / nrow(rwjAll_bil12)
total_ES_12 <- sum(rwjAll_bil12$Spanish, na.rm = TRUE)
ES_percent_12 <- total_ES_12 / nrow(rwjAll_bil12)
total_ER_12 <- sum(rwjAll_bil12$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)

total_EF_11 <- sum(rwjAll_bil11$French, na.rm = TRUE)
EF_percent_11 <- total_EF_11 / nrow(rwjAll_bil11)
total_EA_11 <- sum(rwjAll_bil11$Arabic, na.rm = TRUE)
EA_percent_11 <- total_EA_11 / nrow(rwjAll_bil11)
total_ES_11 <- sum(rwjAll_bil11$Spanish, na.rm = TRUE)
ES_percent_11 <- total_ES_11 / nrow(rwjAll_bil11)
total_ER_11 <- sum(rwjAll_bil11$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.

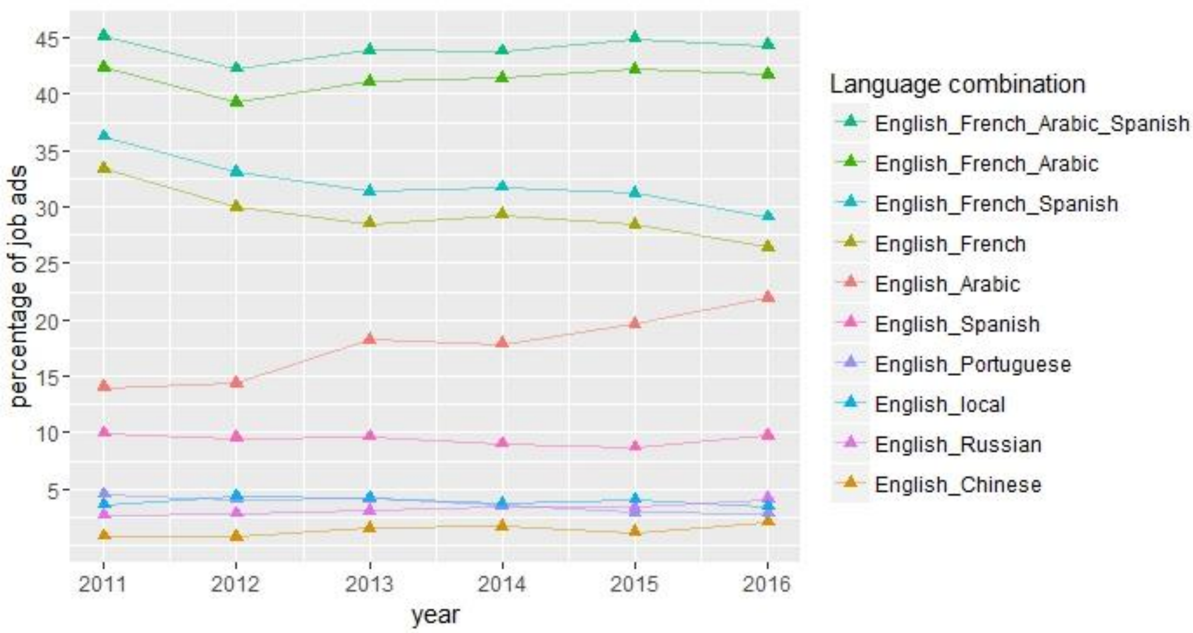
year	English_ French	English_ Arabic	English_ Spanish	English_ Russian	English_ Chinese	English_ Portuguese	English_ local	English_ French_ Arabic	English_ French_ Spanish	English_ French_ Arabic_ Spanish
2011	33.4	14	9.95	2.68	0.81	4.51	3.6	42.39	36.23	45.15
2012	29.95	14.39	9.51	2.81	0.71	4.11	4.37	39.25	33.07	42.29
2013	28.54	18.24	9.65	3.05	1.53	4.12	4.21	41.13	31.4	43.88
2014	29.31	17.86	9.03	3.48	1.66	3.54	3.69	41.42	31.74	43.81
2015	28.41	19.64	8.71	3.37	1.15	2.92	4.06	42.24	31.18	44.93
2016	26.47	21.96	9.74	4.09	2.05	2.83	3.43	41.72	29.12	44.31

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. More than 1 in 4 job ads seek such candidates in 2016, while about 1 in 5 positions requires or prefers English-Arabic speakers. Meanwhile, only around 2 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 roughly 4 in 10 jobs in the aid industry.

- The demand for English-French speakers in global development has declined over the years, from 33.4 percent in 2011 to 26.5 percent as of June 15, 2016. On the other hand, the demand for English-Arabic job candidates is on an upswing, rising from 14 percent in 2011 to nearly 22 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.

<u>One sample t-tests</u>

```
EF_tTest <- t.test(bi_lang_pct$English_French, mu=25)
```

```
Results:
t = 4.6163, df = 5, p-value = 0.005755
alternative hypothesis: true mean is not equal to 25
95 percent confidence interval:
 26.92622 31.76712
sample estimates:
mean of x
 29.34667

EA_tTest <- t.test(bi_lang_pct$English_Arabic, mu=15)
Results:
t = 2.1454, df = 5, p-value = 0.08474
alternative hypothesis: true mean is not equal to 15
95 percent confidence interval:
 14.46853 20.89481
sample estimates:
mean of x
 17.68167

ES_tTest <- t.test(bi_lang_pct$English_Spanish, mu=9)
Results:
t = 2.2546, df = 5, p-value = 0.07385
alternative hypothesis: true mean is not equal to 9
95 percent confidence interval:
 8.939509 9.923825
sample estimates:
mean of x
 9.431667

EC_tTest <- t.test(bi_lang_pct$English_Chinese, mu=1.5)
Results:
t = -0.85522, df = 5, p-value = 0.4315
alternative hypothesis: true mean is not equal to 1.5
95 percent confidence interval:
 0.7722872 1.8643795
sample estimates:
mean of x
 1.318333

ER_tTest <- t.test(bi_lang_pct$English_Russian, mu=3)
Results:
t = 0.56638, df = 5, p-value = 0.5956
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 2.504593 3.775407
sample estimates:
mean of x
     3.14

EP_tTest <- t.test(bi_lang_pct$English_Portuguese, mu=3)
Results:
t = 1.1709, df = 5, p-value = 0.2944
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 2.705147 3.788186
sample estimates:
mean of x
 3.246667

EL_tTest <- t.test(bi_lang_pct$English_local, mu=3)
Results:
t = 12.418, df = 5, p-value = 6.001e-05
alternative hypothesis: true mean is not equal to 2
95 percent confidence interval:
 3.501419 4.285248
```

```
sample estimates:
mean of x
 3.893333

EFA_tTest <- t.test(bi_lang_pct$English_French_Arabic, mu=40)
Results:
t = 2.9241, df = 5, p-value = 0.03286
alternative hypothesis: true mean is not equal to 40
95 percent confidence interval:
 40.16424 42.55243
sample estimates:
mean of x
 41.35833

EFS_tTest <- t.test(bi_lang_pct$English_French_Spanish, mu=30)
Results:
t = 2.1846, df = 5, p-value = 0.08064
alternative hypothesis: true mean is not equal to 30
95 percent confidence interval:
 29.62484 34.62182
sample estimates:
mean of x
 32.12333

EFAS_tTest <- t.test(bi_lang_pct$English_French_Arabic_Spanish, mu=42)
Results:
t = -26.184, df = 5, p-value = 1.518e-06
alternative hypothesis: true mean is not equal to 55
95 percent confidence interval:
 42.98782 45.13551
sample estimates:
mean of x
 44.06167
```

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	1	0
2011	5	9	4
2011	6	22	21
2011	7	31	35
2011	8	57	57
2011	9	30	28
2011	10	52	26
2011	11	42	37
2011	12	50	27
2012	1	37	24
2012	2	74	40
2012	3	38	43
2012	4	38	46
2012	5	17	22
2012	6	69	61
2012	7	29	46
2012	8	32	44
2012	9	46	70
2012	10	40	39
2012	11	42	47
2012	12	30	41
2013	1	54	45
2013	2	41	38
2013	3	48	55
2013	4	52	60
2013	5	56	58
2013	6	65	67
2013	7	58	66
2013	8	50	43
2013	9	73	59
2013	10	43	35

2013	11	23	43
2013	12	45	52
2014	1	65	60
2014	2	35	48
2014	3	36	47
2014	4	59	47
2014	5	40	72
2014	6	40	45
2014	7	52	39
2014	8	44	35
2014	9	60	41
2014	10	50	77
2014	11	57	39
2014	12	47	59
2015	1	34	40
2015	2	35	55
2015	3	39	70
2015	4	40	37
2015	5	37	48
2015	6	0	0
2015	7	72	101
2015	8	45	51
2015	9	30	37
2015	10	50	72
2015	11	24	56
2015	12	48	65
2016	1	39	47
2016	2	57	46
2016	3	44	64
2016	4	38	55
2016	5	42	59
2016	6	29	31

```
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:
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.804396  2.304396
sample estimates:
mean of x mean of y
 41.90625  45.65625
```

The two variables were also tested for correlation:

```
cor(zTestdf$English_Portuguese, zTestdf$English_local)
[1] 0.6737045
```

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.
- English-speaking donor 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", "CÃ´te
d'Ivoire", "Djibouti", "Equatorial Guinea", "Gabon", "Guinea", "Madagascar", "Mali",
"Mauritius", "Morocco", "Niger", "Rwanda", "Senegal", "Togo", "Haiti", "Vanuatu")

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")
```

```
English_donors <- c("United States", "United Kingdom", "Canada", "Ireland", "Australia")

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


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

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

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

ODA_disbursements_French_speaking <-
ODA_disbursements_11_14[ODA_disbursements_11_14$Donor %in% English_donors,]
ODA_disbursements_French_speaking <-
ODA_disbursements_French_speaking[ODA_disbursements_French_speaking$Recipient %in%
French_speaking_countries,]
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")
```

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	8062.1	12099.44	33.4
2012	5518.9	8028.5	29.95
2013	5163.43	8076.64	28.54
2014	5493.06	7506.18	29.31

Year	ODAcommitments_ AS (in US$ millions)	ODAdisbursements_ AS (in US$ millions)	EnglishArabic_ pct

2011	10546.52	14433.07	13.65
2012	9395.02	12167.87	13.98
2013	13244.17	18177.85	17.45
2014	12264.68	14885.27	17.59

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
-0.03026  0.48246 -0.19805 -0.25414

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         20.6389375  1.5771621  13.086   0.0486 *
ODAdisbursements_FS -0.0002845  0.0007544  -0.377   0.7704
ODAcommitments_FS    0.0020136  0.0011949   1.685   0.3409
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5809 on 1 degrees of freedom
Multiple R-squared:  0.9756,      Adjusted R-squared:  0.9267
F-statistic: 19.96 on 2 and 1 DF,  p-value: 0.1563

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

summary(model2)

Residuals:
       1        2        3        4
-0.05004  0.50010 -0.35079 -0.09926

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       2.077e+01  1.163e+00  17.860  0.00312 **
ODAcommitments_FS 1.573e-03  1.885e-04   8.345  0.01406 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.439 on 2 degrees of freedom
Multiple R-squared:  0.9721,      Adjusted R-squared:  0.9581
F-statistic: 69.65 on 1 and 2 DF,  p-value: 0.01406

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

summary(model3)

Residuals:
       1        2        3        4
 0.05901  0.51212 -0.94404  0.37291

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.174e+01  1.989e+00  10.931  0.00827 **
ODAdisbursements_FS 9.588e-04  2.182e-04   4.395  0.04808 *
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.805 on 2 degrees of freedom
Multiple R-squared:  0.9062,      Adjusted R-squared:  0.8592
F-statistic: 19.31 on 1 and 2 DF,  p-value: 0.04808

model4 <- lm(EnglishArabic_pct ~ ODAdisbursements_AS + ODAcommitments_AS, data =
ODAvsEAjobs)
Results:
Residuals:
      1       2       3       4
-0.7250  0.5570  0.3602 -0.1921

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.2500010  3.8531843   0.584    0.664
ODAdisbursements_AS -0.0006818  0.0006371  -1.070    0.478
ODAcommitments_AS    0.0021159  0.0009171   2.307    0.260

Residual standard error: 1.001 on 1 degrees of freedom
Multiple R-squared:  0.9332,      Adjusted R-squared:  0.7996
F-statistic: 6.985 on 2 and 1 DF,  p-value: 0.2585

model5 <- lm(EnglishArabic_pct ~ ODAcommitments_AS, data = ODAvsEAjobs)
Results:
Residuals:
      1       2       3       4
-1.1411  0.6337 -0.1452  0.6527

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       2.4581738  3.9854443   0.617   0.6002
ODAcommitments_AS 0.0012026  0.0003478   3.458   0.0744 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.037 on 2 degrees of freedom
Multiple R-squared:  0.8567,      Adjusted R-squared:  0.7851
F-statistic: 11.96 on 1 and 2 DF,  p-value: 0.07442

model6 <- lm(EnglishArabic_pct ~ ODAdisbursements_AS, data = ODAvsEAjobs)
Results:
Residuals:
      1       2       3       4
-1.7912  0.1526 -0.1200  1.7586

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.8905524  6.2499281   0.942    0.445
ODAdisbursements_AS 0.0006860  0.0004147   1.654    0.240

Residual standard error: 1.78 on 2 degrees of freedom
Multiple R-squared:  0.5777,      Adjusted R-squared:  0.3665
F-statistic: 2.736 on 1 and 2 DF,  p-value: 0.24
```

Correlation among variables were also carried out.

```
cor(ODAvsEFjobs$EnglishFrench_pct, ODAvsEFjobs$ODAdisbursements_FS)
[1] 0.951922
cor(ODAvsEFjobs$EnglishFrench_pct, ODAvsEFjobs$ODAcommitments_FS)
[1] 0.9859438

cor(ODAvsEAjobs$EnglishArabic_pct, ODAvsEAjobs$ODAdisbursements_AS)
[1] 0.7600493
cor(ODAvsEAjobs$EnglishArabic_pct, ODAvsEAjobs$ODAcommitments_AS)
[1] 0.9255846
```

The results indicate that there is a significant relationship between foreign aid commitments and the demand for English-French job candidates, and between foreign aid disbursements and the demand for English-French job candidates. But the relationship between foreign aid commitments and the demand for English-French job candidates is more significant than the relationship between foreign aid disbursements and the demand for English-French job candidates.

In the case of the demand for English-Arabic job candidates, there appears to be no significant relationship with either foreign aid commitments or disbursements, although there is a relatively high correlation between the demand and foreign aid commitments.

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	~ 1 in 4 jobs
Arabic	~ 1 in 5 jobs
Spanish	~ 1 in 10 jobs
Russian	~ 4 in 100 jobs
Local languages	~ 3 in 100 jobs
Portuguese	~ 3 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 (15 percentage points in 2016). Meanwhile, English-French speakers can only see a slight increase (2 percentage points in 2016).

3. The demand for English-Arabic speakers is 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, whether commitments or disbursements, affects trends in the demand for English-French speakers in the aid industry. The same cannot be said for English-Arabic speakers.

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 about 4 in 10 jobs seek qualified candidates who are have knowledge of English, French or Arabic, or a combination of these languages.

3. Universities offering courses focused 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.