

**QE Sandbox: A safe  
space to play with QE  
tools and Method**

**Open Coding and Meaning  
Making: Using QE methods for  
Parsimonious Coding**

Jonathon Sun

# Abstract

Qualitative methods can be daunting because of the sheer number of themes that arise from the data; however, quantitative ethnography (QE) has provided resources to begin making meaning. This workshop will introduce participants to methods to assist in making meaningful qualitative themes, such as a co-occurrence network. First, this workshop will show participants how to clean and import data from other qualitative software like Dedoose. Next, a co-occurrence matrix will be made using the R package/plugin SNAHelper. Finally, based on the selection of themes, an ENA will be built based on the themes from the selected network.

# Goals of this workshop

- Understand the qualitative coding process
- Know qualitative data management systems
- Using R, prepare data for segmentation and ENA
- Make a Network using SNAHelper
- Select codes from Network Analysis for ENA

# Outline of the workshop

- Review Qualitative Research
- Qualitative Research software
- Dedoose --> R
- Network Analysis for code selection
- rENA with the selected codes

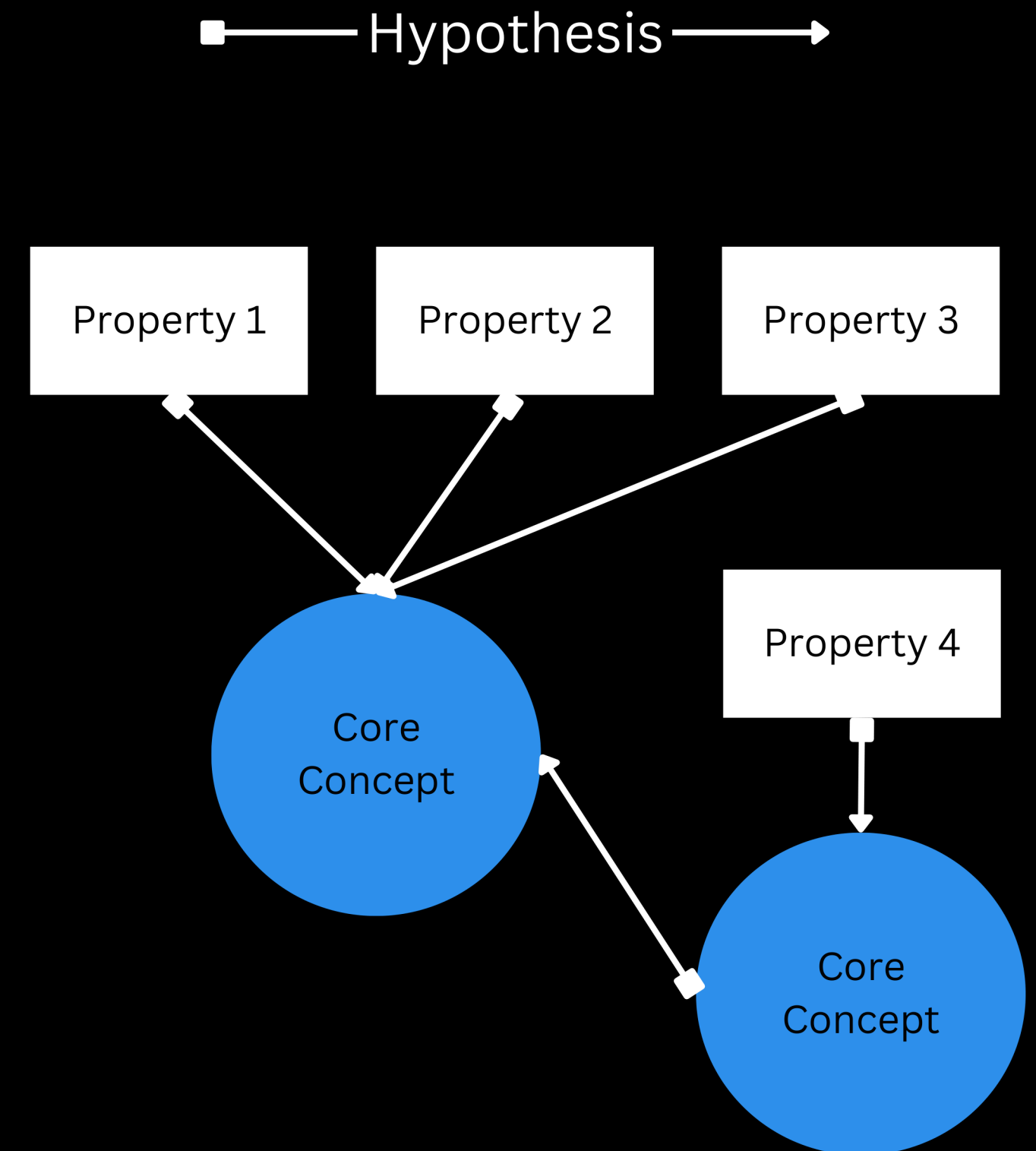
# Qualities of Qualitative Research

***"Qualitative researchers are interested in understanding the meaning people have constructed; that is, how people make sense of their world and the experiences they have in the world."  
(Merriam & Tisdell, 2015, pg. 15)***

- Overall purpose: Achieve an understanding of how people make sense of their lives.
- In qualitative research, we recognize that the researcher is the primary instrument for data collection and analysis.
- Through the process of qualitative research, scholars produce rich/thick descriptions.

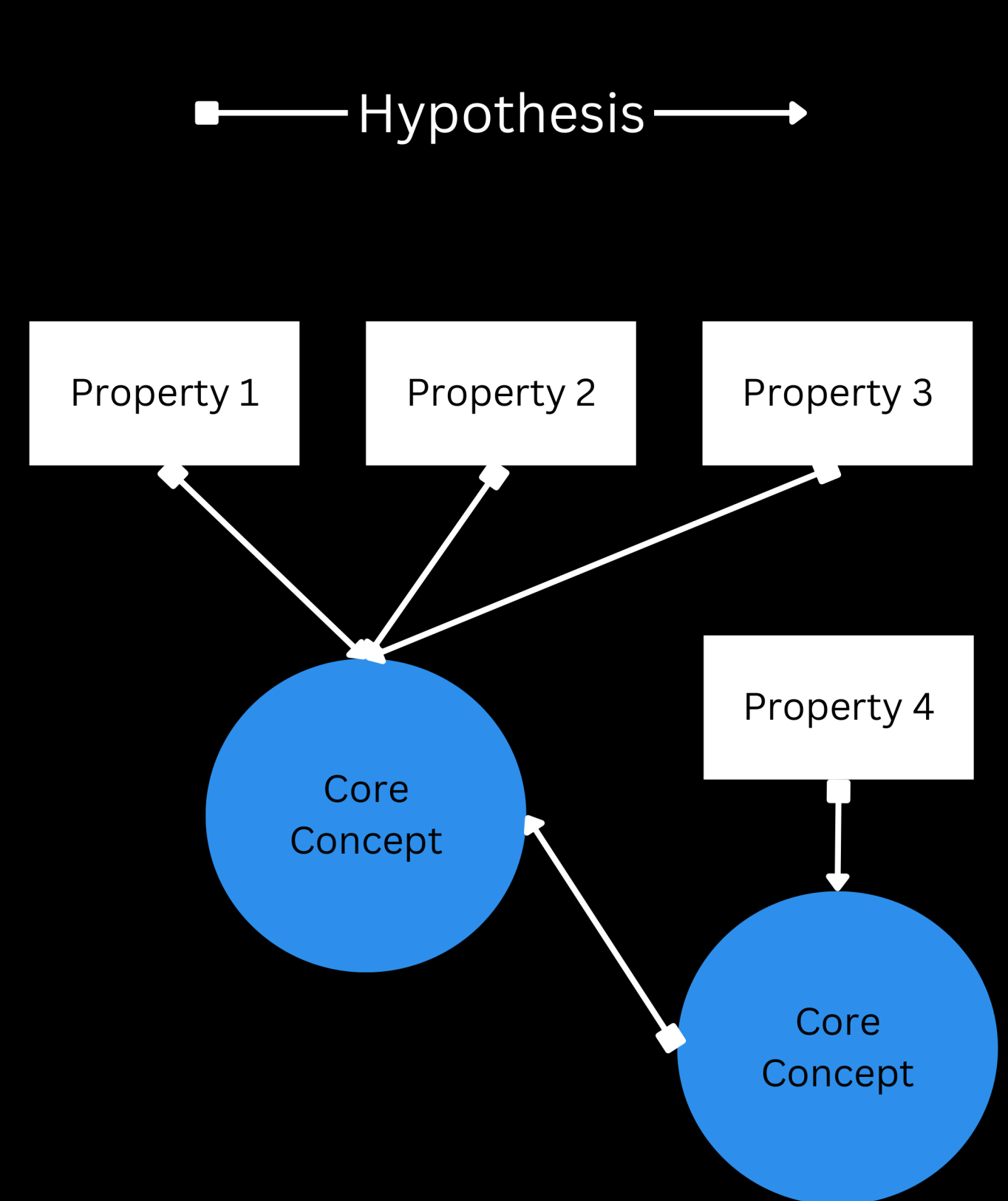
# Qualitative Research Process: Grounded Theory

- Grounded Theory: consists of categories, properties and hypotheses that are conceptual links between and among categories and properties (Merriam & Tisdell, 2015, pg. 228)
- Developing a theory from grounded theory comes from the following: properties, a core category and hypothesis
  - Properties: Concepts that describe a category, consider these like subcodes
  - Core category: The defining aspect of a phenomenon that all other properties or hypothesis are related to
  - Hypothesis: The suggested links between categories and properties

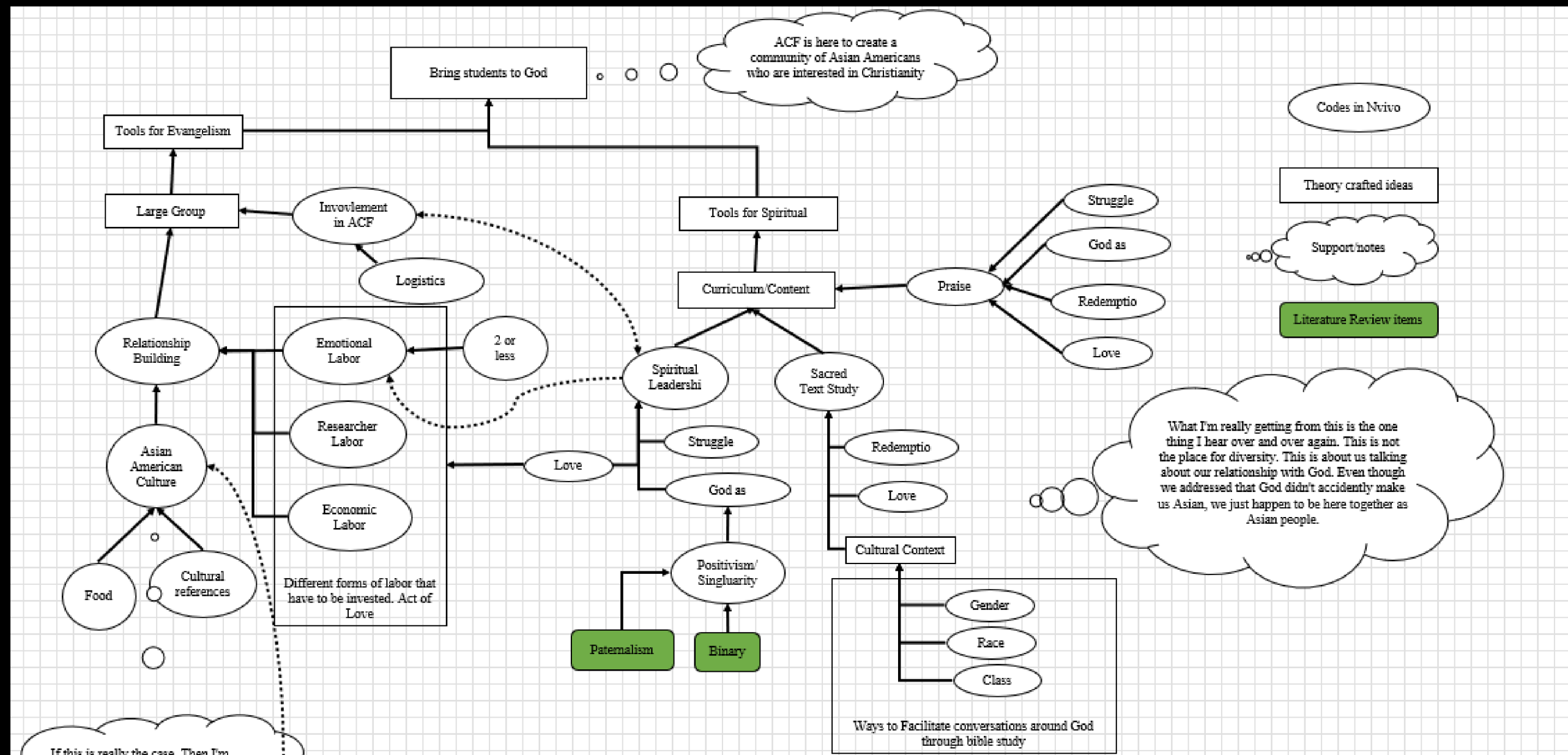


# Qualitative Research Process: Grounded Theory

- Open Coding: First round of coding, tagging any unit of data that may be relevant
- Axial Coding: Relating categories to each other and refining the coding schema
- Selective Coding: propositions or hypothesis are developed



# Qualitative Research Process: Old Example using Excel





Some research hypothesis

Data Collection

Open Coding

Develop Codes

# Qualitative Research Process

Qualitative research is an reflexive process that requires reflection of your own memo taking and flexibility to adjust plans.

THEMES



# Qualitative Data Analysis Software (QDAS)



Qualitative Data Analysis software can be helpful in consolidating your research project.

Each has its own benefits and set of tools to assist in analysis.

However, this study uses Dedoose, because that is what our research team has been using.

# Qualitative Data Analysis Software (QDAS)

segments

codes

CleanText	text	Interview	ID	Code: "Ge	Code: Adu	Code: Adu	Code: Adu	Code: Adu	Code: Adu	Code: Adu	Code: Adu	Code: Bari	Code: Bari	Code: Bari
Empezar en cont	S1 00:00:0	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5151	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
SÃ- [inaudible].	S2 00:00:0	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx		FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Ã¿CuÃ¿intos aÃ¿tos?	S3 00:00:0	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5153	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Ã¿Hace cuÃ¿intos aÃ¿tos viven acÃ¿i?	S1 00:00:1	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5154	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Yo estoy aquÃ- desde el 87.	S3 00:00:1	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5155	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Del 87. Ã¿Y usted?	S4 00:00:1	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5156	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Todos acÃ¿i, casi--	S3 00:00:1	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5157	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Igual.	S5 00:00:1	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5158	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Igual, 87. EstÃ¿i bien.	S4 00:00:1	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5159	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Y yo me mudÃ© aquÃ- en el 2015.	S5 00:00:2	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5160	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Ã¿El 2015? Ã¿Y usted?	S4 00:00:2	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx		FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
No.	S2 00:00:3	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5162	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Ã¿En quÃ© aÃ¿to se mudÃ© aquÃ-?	S5 00:00:3	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5163	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Ã¿En quÃ© aÃ¿to me mudÃ© aquÃ-? En el 2005.	S2 00:00:3	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx		FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
En el 2005. Perfecto.	S4 00:00:3	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5165	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Ã¿Y dÃ¿nde estaban antes de venirse a Filadelfia?	S1 00:00:4	DedooseDoc_Casa Caribe FG1.txt_2023_3_28_1641.docx	5166	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

532

# R Script - Workflow Process

## **Cleaning Interview Documents**

- 1) Import all documents into R as text
- 2) Segment documents using some symbol

## **Prepare Dedoose Data**

- 1) Export data and read into R
- 2) Segment the dedoose data into smaller segmentations then the documents.

## **Merge interview and dedoose data**

- 1) Attach corresponding IDs to segmented portions
- 2) Left\_join codes

# Cleaning Interview Documents: Importing Documents

```
library(officer)
library(tidyverse)
library(dplyr)
```

```
# Importing one document and segmenting by chunks -----
doc <- read_docx(paste0(path, "/", files[1]))
content <- docx_summary(doc) %>%
  separate(text, c("Code", "CleanText"), sep = "S[0-9] [0-9]+:[0-9]+:[0-9]+.[0-9]+",
            remove = FALSE)
```

```
S1 00:00:00.000 Empezar en contándonos hace cuánto viven en Filadelfia.¶
S2 00:00:07.318 Sí [inaudible].¶
S3 00:00:09.276 ¿Cuántos años?¶
S1 00:00:10.221 ¿Hace cuántos años viven acá?¶
S3 00:00:12.005 Yo estoy aquí desde el 87.¶
S4 00:00:14.930 Del 87. ¿Y usted?¶
S3 00:00:18.361 Todos acá, casi--¶
S5 00:00:19.157 Igual.¶
```

doc_index	content_type	style_name	text	Code	CleanText	level	num_id
1	paragraph	NA			NA	NA	NA
2	paragraph	NA	S1 00:00:00.000 Empezar en contándonos hace cuánto viven...		Empezar en contándonos hace cuánto viven en Filadelfia.	NA	NA
3	paragraph	NA			NA	NA	NA
4	paragraph	NA	S2 00:00:07.318 Sí [inaudible].		Sí [inaudible].	NA	NA
5	paragraph	NA			NA	NA	NA
6	paragraph	NA	S3 00:00:09.276 ¿Cuántos años?		¿Cuántos años?	NA	NA
7	paragraph	NA			NA	NA	NA
8	paragraph	NA	S1 00:00:10.221 ¿Hace cuántos años viven acá?		¿Hace cuántos años viven acá?	NA	NA
9	paragraph	NA			NA	NA	NA
10	paragraph	NA	S3 00:00:12.005 Yo estoy aquí desde el 87.		Yo estoy aquí desde el 87.	NA	NA
11	paragraph	NA			NA	NA	NA
12	paragraph	NA	S4 00:00:14.930 Del 87. ¿Y usted?		Del 87. ¿Y usted?	NA	NA
13	paragraph	NA			NA	NA	NA
14	paragraph	NA	S3 00:00:18.361 Todos acá, casi--		Todos acá, casi--	NA	NA
15	paragraph	NA			NA	NA	NA
16	paragraph	NA	S5 00:00:19.157 Igual.		Igual.	NA	NA
17	paragraph	NA			NA	NA	NA
18	paragraph	NA	S4 00:00:19.461 Igual, 87. Está bien.		Igual, 87. Está bien.	NA	NA
19	paragraph	NA			NA	NA	NA
20	paragraph	NA	S5 00:00:22.200 Y yo me mudé aquí en el 2015.		Y yo me mudé aquí en el 2015.	NA	NA

# Cleaning Interview Documents: Cleaning Text

```
# Pulls Participant and time stamp
```

```
# str_extract(content$text, "S[0-9] [0-9]+:[0-9]+:[0-9]+.[0-9]+") %>%  
# as.data.frame()
```

```
# Extract Participant
```

```
content <- cbind(content, str_extract(content$text, "S[0-9]")) %>%  
  as.data.frame() %>%  
  rename("Participant" = ".")
```

```
# Extract timestamp
```

```
content <- cbind(content, str_extract(content$text, "[0-9]+:[0-9]+:[0-9]+.[0-9]+")) %>%  
  as.data.frame() %>%  
  rename("TimeStamp" = ".")
```

```
content <- content %>%
```

```
  select(Participant, TimeStamp, CleanText, text) %>%  
  na.omit()
```

	Participant	TimeStamp	CleanText	text
2	S1	00:00:00.000	Empezar en contándonos hace cuánto viven en Filadelfia.	S1 00:00:00.000 Empezar en contándonos hace cuánto viven...
4	S2	00:00:07.318	Sí [inaudible].	S2 00:00:07.318 Sí [inaudible].
6	S3	00:00:09.276	¿Cuántos años?	S3 00:00:09.276 ¿Cuántos años?
8	S1	00:00:10.221	¿Hace cuántos años viven acá?	S1 00:00:10.221 ¿Hace cuántos años viven acá?
10	S3	00:00:12.005	Yo estoy aquí desde el 87.	S3 00:00:12.005 Yo estoy aquí desde el 87.
12	S4	00:00:14.930	Del 87. ¿Y usted?	S4 00:00:14.930 Del 87. ¿Y usted?
14	S3	00:00:18.361	Todos acá, casi--	S3 00:00:18.361 Todos acá, casi--
16	S5	00:00:19.157	Igual.	S5 00:00:19.157 Igual.
18	S4	00:00:19.461	Igual, 87. Está bien.	S4 00:00:19.461 Igual, 87. Está bien.
20	S5	00:00:22.200	Y yo me mudé aquí en el 2015.	S5 00:00:22.200 Y yo me mudé aquí en el 2015.
22	S4	00:00:26.680	¿El 2015? ¿Y usted?	S4 00:00:26.680 ¿El 2015? ¿Y usted?
24	S2	00:00:30.601	No.	S2 00:00:30.601 No.
26	S5	00:00:30.994	¿En qué año se mudó aquí?	S5 00:00:30.994 ¿En qué año se mudó aquí?
28	S2	00:00:32.968	¿En qué año me mudé aquí? En el 2005.	S2 00:00:32.968 ¿En qué año me mudé aquí? En el 2005.
30	S4	00:00:38.307	En el 2005. Perfecto.	S4 00:00:38.307 En el 2005. Perfecto.
32	S1	00:00:40.699	¿Y dónde estaban antes de venirse a Filadelfia?	S1 00:00:40.699 ¿Y dónde estaban antes de venirse a Filadel...
34	S5	00:00:43.947	Yo estaba en Connecticut. Vine aquí a poner una bodega a...	S5 00:00:43.947 Yo estaba en Connecticut. Vine aquí a poner...
36	S4	00:00:53.309	¿Y antes de Connecticut dónde estaba?	S4 00:00:53.309 ¿Y antes de Connecticut dónde estaba?
38	S5	00:00:57.738	En Nueva York.	S5 00:00:57.738 En Nueva York.
40	S4	00:00:59.827	¿Y antes de eso?	S4 00:00:59.827 ¿Y antes de eso?
42	S5	00:01:01.457	¿Antes de Nueva York? En Nueva York fue que yo vine. De a...	S5 00:01:01.457 ¿Antes de Nueva York? En Nueva York fue q...
44	S4	00:01:10.915	¿Y antes de Nueva York? ¿De dónde emigró?	S4 00:01:10.915 ¿Y antes de Nueva York? ¿De dónde emigró?
46	S5	00:01:13.488	En Santo Domingo.	S5 00:01:13.488 En Santo Domingo.
48	S4	00:01:15.038	¿Y qué parte de Santo Domingo?	S4 00:01:15.038 ¿Y qué parte de Santo Domingo?

# Cleaning Interview Documents: Make it a Function!

Custom functions helps writing more efficient code.

It's easier to make changes to overall code, and easier to keep track of because, less copy pasting

```
Segmenting <- function(doc){
  content <- docx_summary(doc) %>%
    separate(text, c("Code", "CleanText"), sep = "S[0-9] [0-9]+:[0-9]+:[0-9]+.[0-9]+",
             remove = FALSE)

  #Pulls Participant and time stamp
  #str_extract(content$text, "S[0-9] [0-9]+:[0-9]+:[0-9]+.[0-9]+") %>%
  # as.data.frame()

  #Extract Participant
  content <- cbind(content, str_extract(content$text, "S[0-9]") %>%
                  as.data.frame()) %>%
    rename("Participant" = ".")

  #Extract timestamp
  content <- cbind(content, str_extract(content$text, "[0-9]+:[0-9]+:[0-9]+.[0-9]+") %>%
                  as.data.frame()) %>%
    rename("TimeStamp" = ".")

  content <- content %>%
    select(Participant, TimeStamp, CleanText, text) %>%
    na.omit()
  content <<- content
}
```

# LOOP IT

For loops are not the fastest and most efficient code. They are easy to understand and use.

```
#Looping and importing all documents -----  
for(i in 1:length(files)) {  
  print(paste0(i, " out of ", length(files)))  
  if(i == 1) {  
    doc <- read_docx(paste0(path,"/",files[i]))  
    Segmenting(doc = doc)  
    All <- content %>%  
      mutate(Interview = files[i])  
  } else {  
    doc <- read_docx(paste0(path,"/",files[i]))  
    Segmenting(doc = doc)  
    All <- rbind(All,content %>%  
      mutate(Interview = files[i]))  
  }  
}
```



# Preparing Dedoose Data

```
library(readxl)
```

```
path <- "Data/FGExcerptsTransit"
```

```
list.files(path)
```

```
TransitExcerpts <- read_excel(paste0(path,"/",list.files(path)[1])) %>%  
  filter(!is.na(`Excerpt Copy`)) %>%  
  mutate(ID = row_number())
```

	Media Title	Excerpt Range	Excerpt Creator	Excerpt Date	Excerpt Copy	Resource Creator	Codes Applied Combined	Code: "Getting out" of the Ghetto Applied
1	Lehigh HACE staff FG2.txt	2141-2667	HED-Sun	2/8/2023	S2 00:02:24.142 I've lived in the same house, same block, m...	AstridPickenpack	North Philly El Norte	False
2	Lehigh HACE staff FG2.txt	2668-2822	HED-Sun	2/8/2023	S1 00:02:42.723 Okay, excellent. How about you? S3 00:02:4...	AstridPickenpack	North Philly El Norte	False
3	Lehigh HACE staff FG2.txt	2823-3924	HED-Sun	2/9/2023	S1 00:02:49.821 Okay. Okay. And how about you? S4 00:02:5...	AstridPickenpack	Living in Philadelphia	False
4	Lehigh HACE staff FG2.txt	3926-4994	HED-Sun	2/9/2023	S1 00:03:48.211 Yeah. Okay. So how do you personally ident...	AstridPickenpack	Identity	False
5	Lehigh HACE staff FG2.txt	4995-6058	HED-Sun	2/9/2023	S4 00:04:53.862 This will be one year. One year at HACE. S1 ...	AstridPickenpack	Employment	False
6	Lehigh HACE staff FG2.txt	5632-5765	HED-Sun	7/15/2022	S3 00:05:22.075 HACE was my first job. I would only do sum...	AstridPickenpack	Community Agencies	False
7	Lehigh HACE staff FG2.txt	6060-7563	HED-Sun	2/9/2023	So tell me a little bit. It seems like you're all kind of close to ...	AstridPickenpack	Primary/Middle/High Schools, Living in Philadelphia	False
8	Lehigh HACE staff FG2.txt	7564-12393	HED-Sun	2/9/2023	S1 00:07:24.724 Okay. Okay. How about you? S3 00:07:27.41...	AstridPickenpack	Primary/Middle/High Schools	False
9	Lehigh HACE staff FG2.txt	7753-8743	HED-Sun	2/9/2023	S3 00:07:38.087 Middle school, yes. I do remember it. But w...	AstridPickenpack	School	False
10	Lehigh HACE staff FG2.txt	12394-13230	HED-Sun	2/9/2023	S1 00:12:09.964 Okay. All right. I understand. Okay. How ab...	AstridPickenpack	Primary/Middle/High Schools, Predisposition	False

*more codes ...*  
→

# Preparing Dedoose Data

```
str_split(TransitExcerpts$`Excerpt Copy`[i], "\r") %>%  
  unlist() %>%  
  as.data.frame() %>%  
  dplyr::rename("Excerpt" = ".") %>%  
  filter(!Excerpt == "\n")
```

What we need R to do	What R does
1) From the coded Dedoose data, split all the data by /r (line break).	R does this with the str_split function, and gives us a list of all the segments where \r is present.
2) Unlist this and make it a data frame	This is just for me, it's easier to work in dataframes than it is a list.
3) We then change the name of the newly made column from . --> "Excerpt"	Gives us a datatable with a renamed column
4) Filter out any empty lines	Removes lines that are empty

# Preparing Dedoose Data

```
[1] "732 of 1243"  
[1] "733 of 1243"  
[1] "734 of 1243"  
[1] "735 of 1243"  
[1] "736 of 1243"  
[1] "737 of 1243"  
[1] "738 of 1243"  
[1] "739 of 1243"
```

This gives us a new dedoose code book where each chunk of coded material has been segmented into smaller portions because it's split at "/r"

We also retained all the codes corresponding to each line. Because we separated these into smaller chunks than the segmented data we can then line up any text that

	Excerpt	Code: "Getting out" of the Ghetto Applied	Code: Adult Education Applied	Code: Adult Education\Community Agencies Applied
1	S2 00:02:24.142 I've lived in the same house, same block, m...	False	False	False
2	S1 00:02:26.956 Okay, and where in Philadelphia have you li...	False	False	False
3	S2 00:02:29.819 I live in Kensington.	False	False	False
4	S1 00:02:31.283 In Kensington.	False	False	False
5	S2 00:02:31.529 I can see Kensington Ave from my house.	False	False	False
6	S1 00:02:33.695 You can see what?	False	False	False

# Merging the sets

```
#Loops through all excerpts and assigns matching code
for(i in 1:nrow(ExcerptsAll)) {
  print(paste(i," of ", nrow(ExcerptsAll)))
  check <- str_remove(ExcerptsAll$Excerpt[i],"\n")
  if(grepl("\\[risas",check) == TRUE){
    check <- str_remove(check,"\\[risas")

    ID <- ExcerptsAll$ID[i]
    All <- All %>%
      mutate(CodeID = ifelse(grepl(check,text),ID,CodeID))
  } else if (grepl("\\[Nunca, ¿ve?",check) == TRUE) {
    check <- str_remove(check,"\\[Nunca, ¿ve?")

    ID <- ExcerptsAll$ID[i]
    All <- All %>%
      mutate(CodeID = ifelse(grepl(check,text),ID,CodeID))
  } else if (grepl("\\[inaudible",check) == TRUE) {
    check <- str_remove(check,"\\[inaudible")

    ID <- ExcerptsAll$ID[i]
    All <- All %>%
      mutate(CodeID = ifelse(grepl(check,text),ID,CodeID))
  } else {
    ID <- ExcerptsAll$ID[i]
    All <- All %>%
      mutate(CodeID = ifelse(grepl(check,text),ID,CodeID))
  }
}
```

At this point, I have assigned a unique number to both my Dedoose data set and my segmented text.

My goal at this point is check what text is shared between the two datasets, and then assign the segmented text a unique number from the coded dedoose dataset.

The script

- 1) Removes the text \n from the text and saves it as "check"
- 2) Using grepl, R checks if the phrase is found in any of the segments from the segmented text. If there is a match, the unique ID from the coded text is attached to the segmented transcript text.
- 3) There are some words that R did not like, and I don't know why. I created special cases for them.



# Aggregating Subcodes into Primary Codes

Dedoose does not automatically count subcodes to primary code frequencies.

This study analysis relies on the primary codes, because this is the open coding portion and there are a lot of codes. By aggregating subcodes, we can change the granularity

```
#Identify all primary codes and separate into vector  
PrimaryCodes <- cbind(colnames(All %>%  
  select(starts_with("Code"))),  
  grepl("\\\\",colnames(All %>%  
    select(starts_with("Code"))))) %>%  
as_tibble() %>%  
filter(V2 == FALSE) %>%  
mutate(V1 = str_replace_all(V1,"\\\\", "_"),  
  V1 = str_remove(V1, "Applied"),  
  V1 = str_remove(V1, "Code:")) %>%  
select(V1)
```

```

for(i in 1:nrow(PrimaryCodes)) {
  print(paste(i," of ", nrow(PrimaryCodes)))

  loop <- AllRename %>%
    dplyr::select(ID, contains(PrimaryCodes$V1[i]))

  if(length(loop) < 3) next

  #If columns 3:nrow(loop) = TRUE. Any of them, THEN
  column 2 = TRUE
  #for row 1, select a range of columns and use paste and
  make them a string
  loop <- loop %>%
    unite(paste,3:length(loop)) %>%
    mutate(Hold = ifelse(grepl("True",paste),"True",
"False")) %>%
    select(ID, Hold)

  #Change the ColName
  ColName <- paste0(str_squish(paste0("Code:
",PrimaryCodes$V1[i]," Applied")))

  #Identifies the ColName we are looking to change
  AllRename[,ColName] <- loop$Hold
}

```

## Aggregating Subcodes into Primary Codes

Loop subsets all codes related to primary codes, then determines if any of the subcodes are true, but creating a string that includes all values from columns. This then creates a new column called Paste which is then passed into grepl to determine if "True" is in the string. If "True" is in the string then the new column HOLD is "TRUE"

# Co-Occurrence Matrix

- 1) Convert True/False values to logical value
- 2) Select only primary codes
- 3) Create co-occurrence matrix with script from Stackoverflow

	Code...Getting.out..of.the.Ghetto.Applied	Code..Adult.Education.Applied	Code..Barriers.to.Education.Applied
Code...Getting.out.of.the.Ghetto.Applied	0	0	0
Code..Adult.Education.Applied	0	0	98
Code..Barriers.to.Education.Applied	0	98	0
Code..Calle.Education.Applied	0	0	0
Code..College.Access.Applied	28	177	102
Code..Dominican.Republic.Applied	0	23	0
Code..Employment.Applied	0	57	43
Code..Family.Migration.Applied	6	29	3
Code..Health.Challenges.Applied	0	42	34
Code..Homelessness.Applied	0	3	0
Code..Housing.conflicts.Applied	25	0	0
Code..Identity.Applied	3	0	0
Code..Immigration.Applied	0	0	0
Code..Incarceration.Applied	7	4	0
Code..Living.in.Philadelphia.Applied	24	15	0
Code..Mobility.Applied	14	5	0

```
CoOcc1 <- All %>%
  select(ID, starts_with("Code:"))

PrimaryCodesCol <- colnames(CoOcc1) %>%
  as_tibble() %>%
  filter(!grepl("QQQ",value))
```

```
CoOcc1 <- CoOcc1[PrimaryCodesCol$value]
```

```
#changing all columns to upper case to then transform to binary
CoOcc1 <- data.frame(lapply(CoOcc1, function(v) {
  if (is.character(v)) return(toupper(v))
  else return(v)
}))
```

```
#Changing from character to Binary
CoOcc1 <- CoOcc1 %>%
  mutate_if(is.character, as.logical)
```

```
#Removing ID because it is not needed for the analysis and running the co-occurrence matrix
```

```
# https://stackoverflow.com/questions/10622730/create-a-co-occurrence-matrix-from-dummy-coded-observations
```

```
X <- as.matrix(CoOcc1 %>%
  select(!ID))
```

```
out <- crossprod(X)
```

```
diag(out) <- 0
```

```
out <- as.data.frame(out)
```



# Co-Occurrence Matrix to Network Analysis

```
FrequencyCoded <- rowSums(out) %>%  
  as_tibble() %>%  
  cbind(colnames(out)) %>%  
  rename(Code = `colnames(out)`)
```

```
CoOccur <- rownames_to_column(out, "V1") %>%  
  pivot_longer(starts_with("Code"),  
    names_to = "V2",  
    values_to = "CoOccurrence")
```

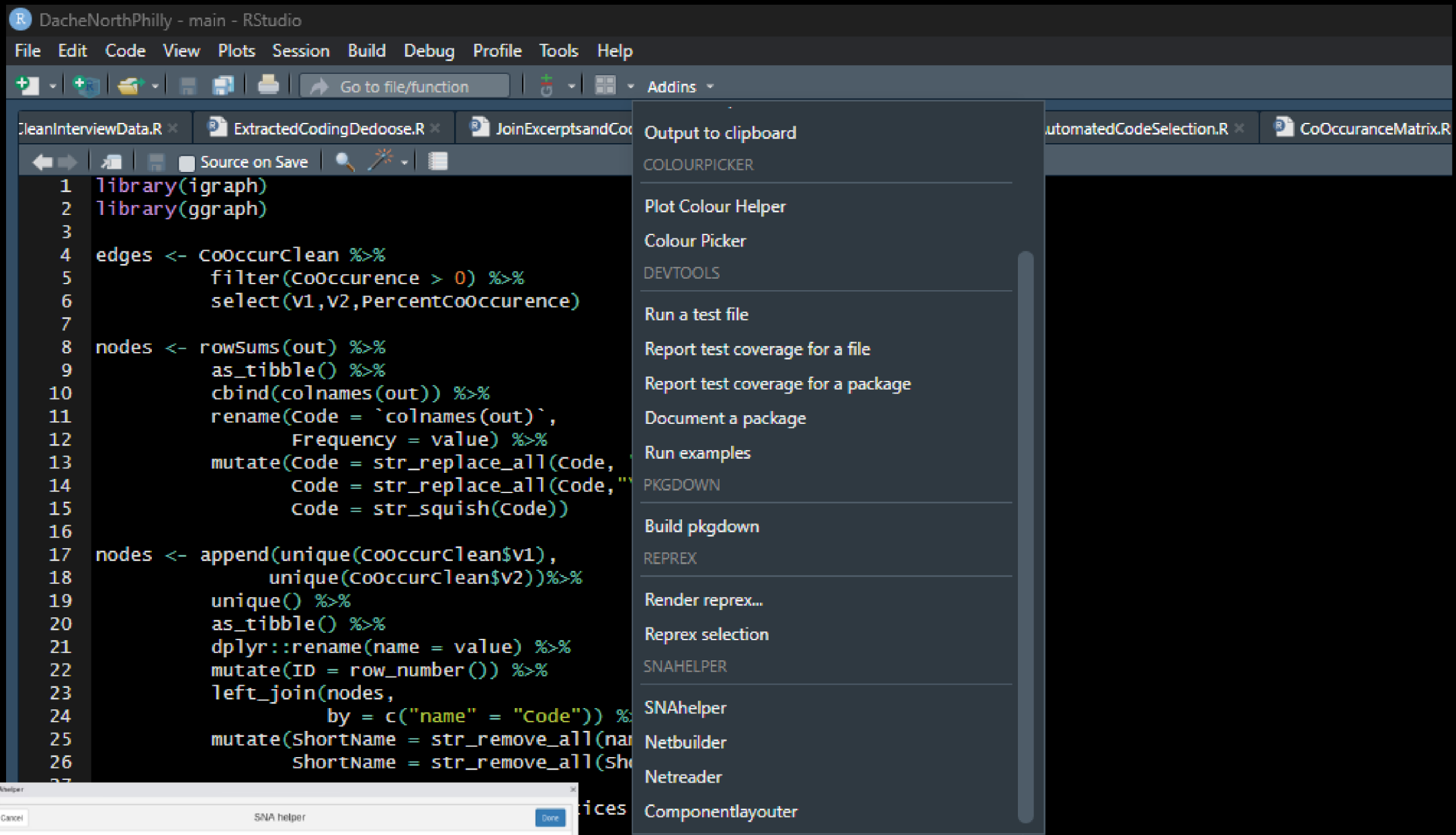
```
CoOccur <- left_join(CoOccur,  
  FrequencyCoded,  
  by = c("V1" = "Code")) %>%  
  rename(Total = value)
```

V1	V2	CoOccurrence	Total
Code...Getting.out.of.the.Ghetto.Applied	Code..Adult.Education.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Barriers.to.Education.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Calle.Education.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..College.Access.Applied	28	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Dominican.Republic.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Employment.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Family.Migration.Applied	6	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Health.Challenges.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Homelessness.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Housing.conflicts.Applied	25	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Identity.Applied	3	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Immigration.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Incarceration.Applied	7	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Living.in.Philadelphia.Applied	24	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Mobility.Applied	14	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Motivators.for.Education.Applied	0	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Neighborhood.Boundaries.Applied	15	220
Code...Getting.out.of.the.Ghetto.Applied	Code..Opiod.Crises.Substance.Use.Applied	3	220

# Network Analysis

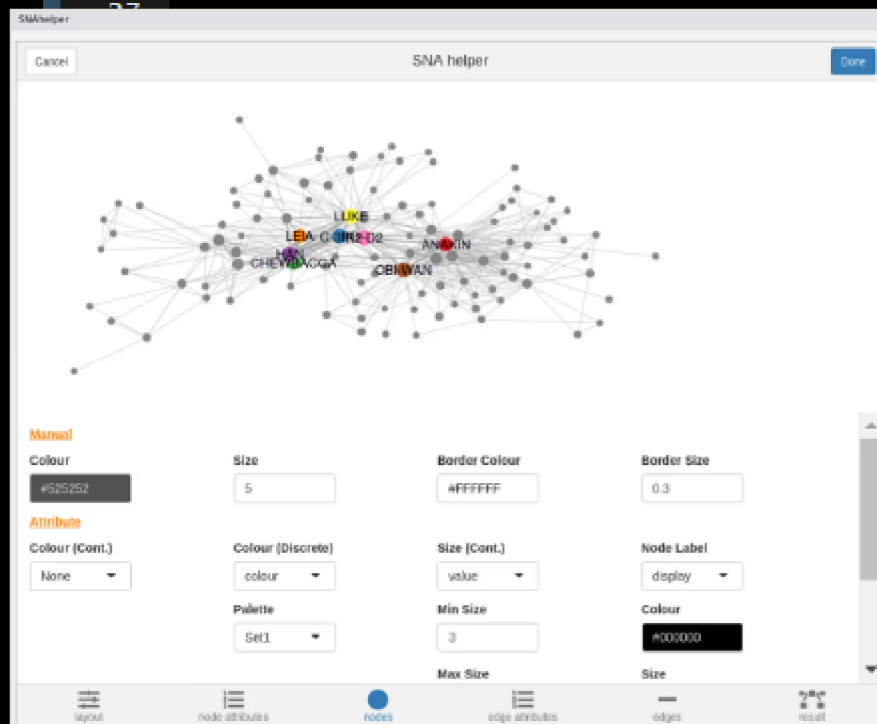
## Package: igraph and ggraph Plugin: SNAhelper

- 1) Specify your edge and nodes
- 2) create a "Graph from data frame"
- 3) highlight the object and open "SNAhelper"



```
1 library(igraph)
2 library(ggraph)
3
4 edges <- CooccurClean %>%
5   filter(CoOccurrence > 0) %>%
6   select(v1,v2,PercentCoOccurrence)
7
8 nodes <- rowSums(out) %>%
9   as_tibble() %>%
10  cbind(colnames(out)) %>%
11  rename(Code = `colnames(out)` ,
12        Frequency = value) %>%
13  mutate(Code = str_replace_all(Code,
14    Code = str_replace_all(code,
15    Code = str_squish(code))
16
17 nodes <- append(unique(CooccurClean$v1),
18   unique(CooccurClean$v2))%>%
19  unique() %>%
20  as_tibble() %>%
21  dplyr::rename(name = value) %>%
22  mutate(ID = row_number()) %>%
23  left_join(nodes,
24    by = c("name" = "Code")) %>%
25  mutate(shortName = str_remove_all(name,
26    shortName = str_remove_all(shortName,
27
```

The screenshot shows the RStudio interface with the Addins menu open. The menu includes options like 'Output to clipboard', 'COLOURPICKER', 'Plot Colour Helper', 'Colour Picker', 'DEVTOOLS', 'Run a test file', 'Report test coverage for a file', 'Report test coverage for a package', 'Document a package', 'Run examples', 'PKGDOWN', 'Build pkgdown', 'REPREX', 'Render reprex...', 'Reprex selection', 'SNAHELPER', 'SNAhelper', 'Netbuilder', 'Netreader', and 'Componentlayerouter'.



### schochastics/snahelper: Rstudio Addin for Network Analysis and Visualization

Rstudio Addin for Network Analysis and Visualization - GitHub - schochastics/snahelper: Rstudio Addin for Network Analysis and Visualization

README.md

## snahelper

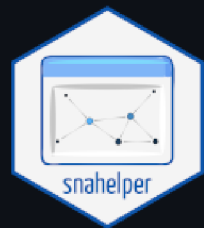
R-CMD-check **passing** CRAN **1.4.1** downloads **278/month**

**snahelper** provides a set RStudio Addin for social network analysis. The main addin is the **SNAhelper** which provides a simple GUI to do common network analytic tasks and visualize a network with **ggraph**.

The second addin, called **Netbuilder** allows you to quickly build small networks with a small "canvas" to draw on. The network can be exported as an **igraph** object at the end of the session by clicking on "Done".

The third addin **Netreader** is meant to facilitated the import of raw network data. It provides a GUI to easily read network and attribute data and combine them to an **igraph** object. The underlying code of the import procedure is shown at the end. This should help users to learn importing data themselves.

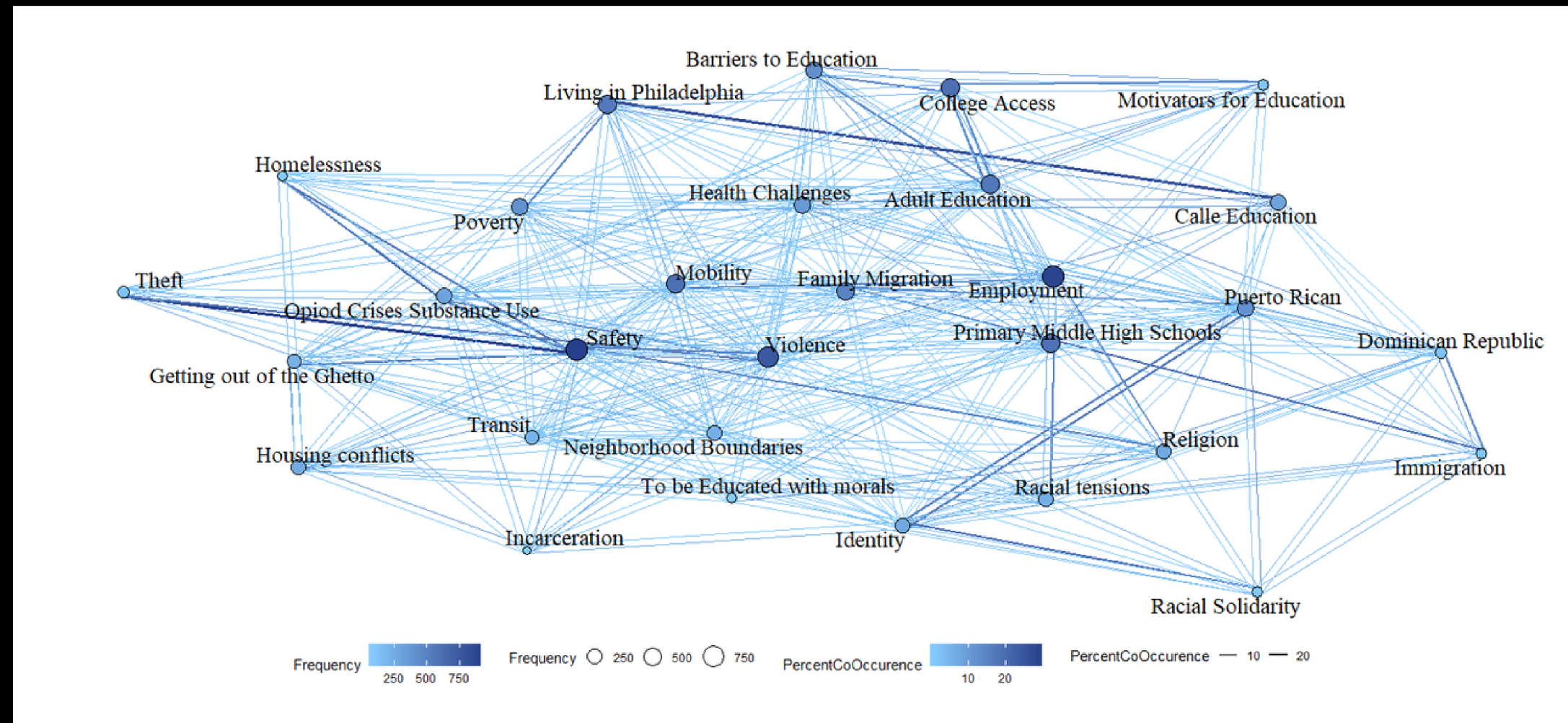
The fourth addin **Componentlayerouter** allows to layout networks with several components manually by placing them on an empty canvas. Components can also be rotated and resized. After finishing the session, the layout is saved as x and y vertex attributes.



## #Network analysis with weights as Co-Occurrence percentages

```
x <- c(-1.0762, 0.3477, -0.0124, 0.9355, 0.2663, 1.2693, 0.4759, 0.0522, -0.0367, -1.1007, -1.0666, 0.1683, 1.3529, -0.6002, -0.4357, -0.2949, 0.9068, -0.2156, -0.7706, -0.6139, 0.4713, 0.87, 0.895, 0.4619, 0.7032, -0.4983, -1.4261, -0.1825, -0.5913, -0.1067)  
y <- c(-0.1618, 0.6521, 1.1774, 0.5724, 1.0969, -0.1216, 0.2317, 0.1631, 0.5598, 0.6907, -0.6484, -0.9176, -0.5857, -1.0308, 1.022, 0.1922, 1.1094, -0.4925, 0.1385, 0.5487, -0.0809, 0.0809, -1.2238, -0.7989, -0.5782, -0.1089, 0.1587, -0.7917, -0.5096, -0.143)
```

```
ggraph(g, layout = "manual", x = x, y = y) +  
  geom_edge_parallel0(aes(width = PercentCoOccurence, colour = PercentCoOccurence), edge_alpha = 1) +  
  scale_edge_colour_gradient(low = "#87CEFF",  
    high = "#27408B") +  
  scale_edge_width(range = c(0.3, 1.2)) +  
  geom_node_point(aes(fill = Frequency, size = Frequency), colour = "#000000", shape = 21, stroke = 0.3) +  
  scale_fill_gradient(low = "#87CEFF", high = "#27408B") +  
  scale_size(range = c(3, 8)) +  
  geom_node_text(aes(label = ShortName), colour = "#000000", size = 6,  
    family = "serif", repel = TRUE, segment.alpha = 0) +  
  theme_graph() +  
  theme(legend.position = "bottom")
```



# Selecting Themes

**QE provides tools to analyze your data and reveal relationships that may not have been clear before; however, as researchers, we are still responsible for interpretation**

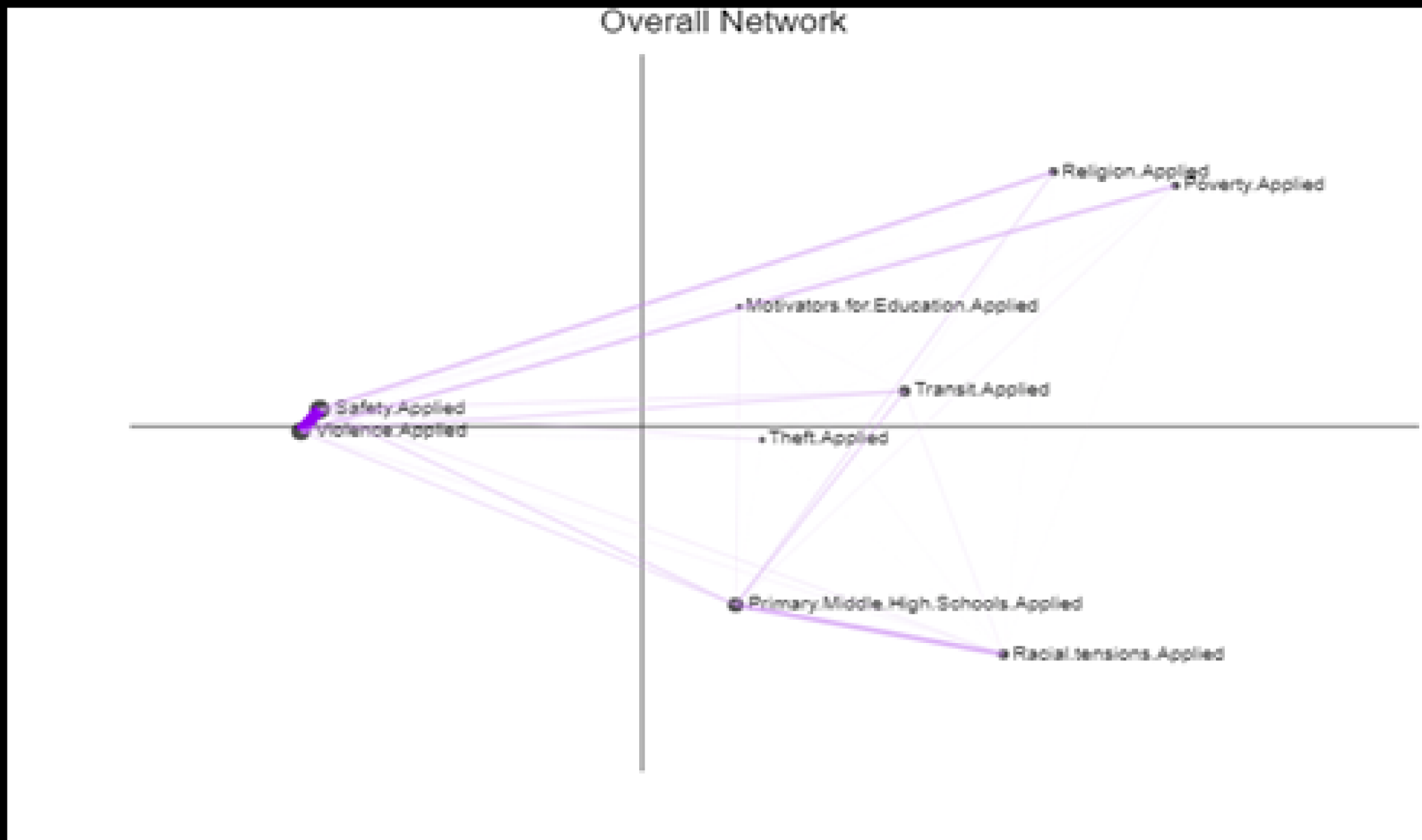
Seven pairs with the highest co-occurring percentages were selected as codes:

- (1) safety,
- (2) violence,
- (3) religion,
- (4) poverty,
- (5) motivators for education,
- (6) transit,
- (7) theft,
- (8) primary, middle high school,
- (9) racial tensions.

As a research team, these were recurring themes that we noticed. Network analysis supported the ideas that we had in mind and provided other codes to consider

# Network Analysis to Epistemic Network Analysis

ENA differs from network analysis because it shows the relationship based on the discourse. The co-occurrence network uses the coded text but does not consider the text's relationship within the discourse.



- Relationship between safety and violence.
- While transit as a code may not have strong relationships with other codes and is semantically distant, the location of the point suggests that transit has some relationship with other codes.
- This code selection also raises an interesting theme about the relationship between religion and poverty, which the research team has not explored.

Some research hypothesis



Data Collection

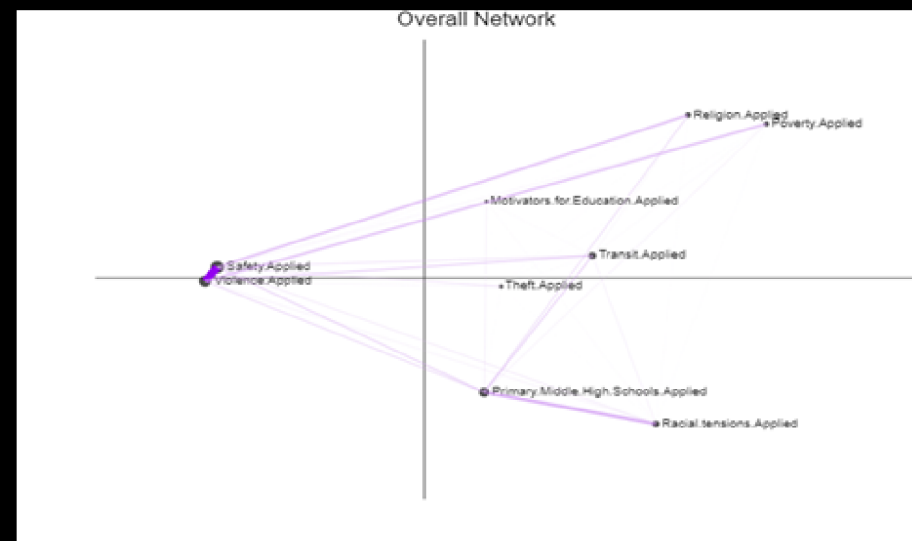
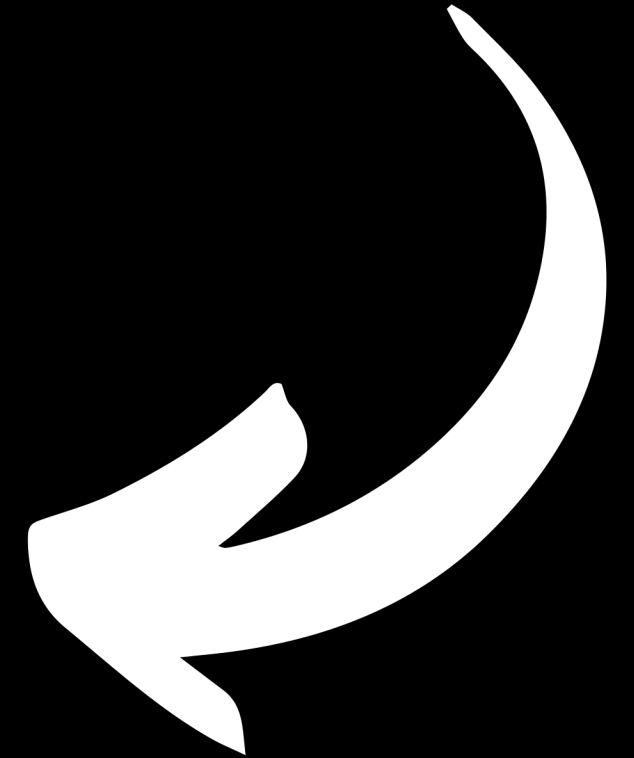
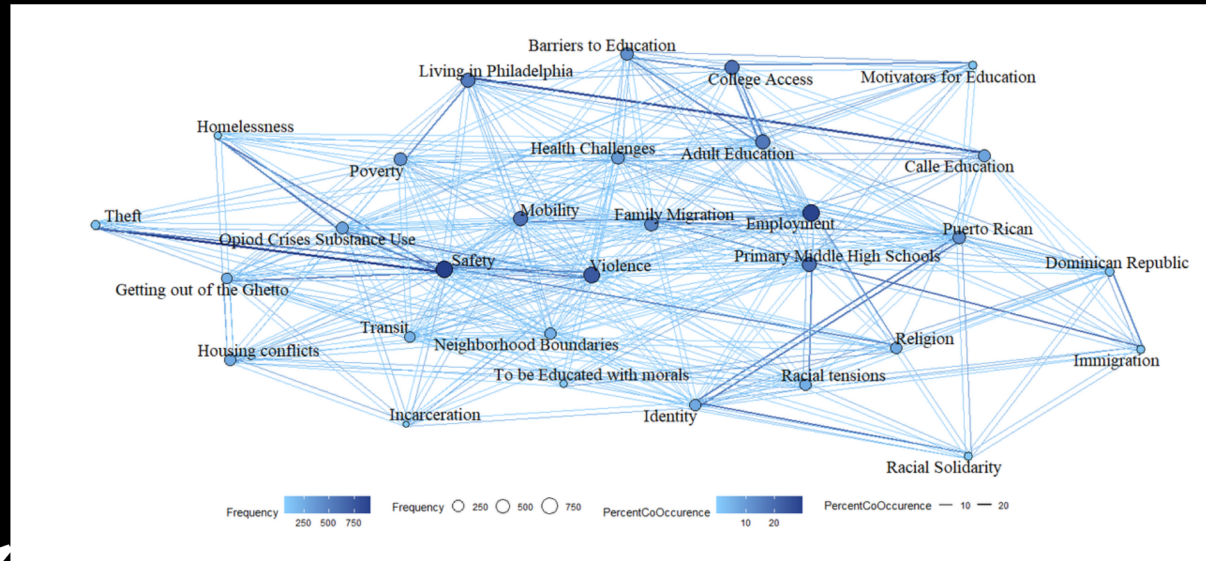
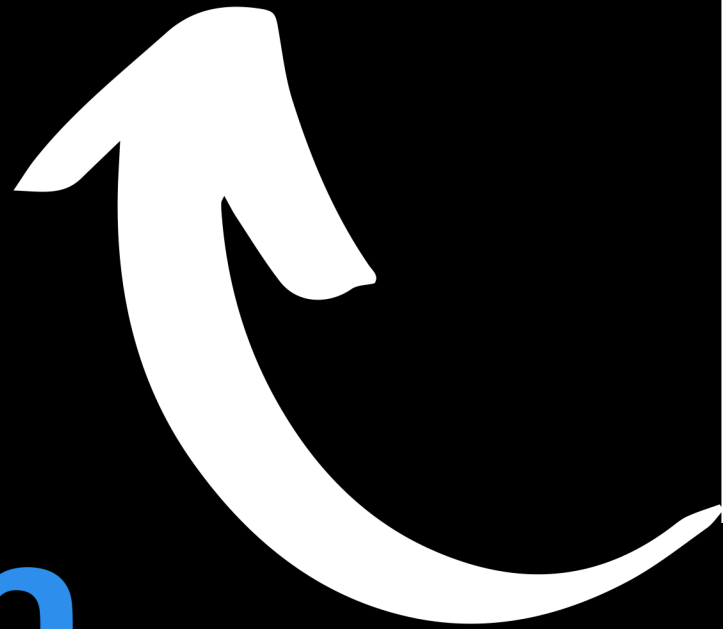


Open Coding



Develop Codes

# QE Research Process



THEMES

