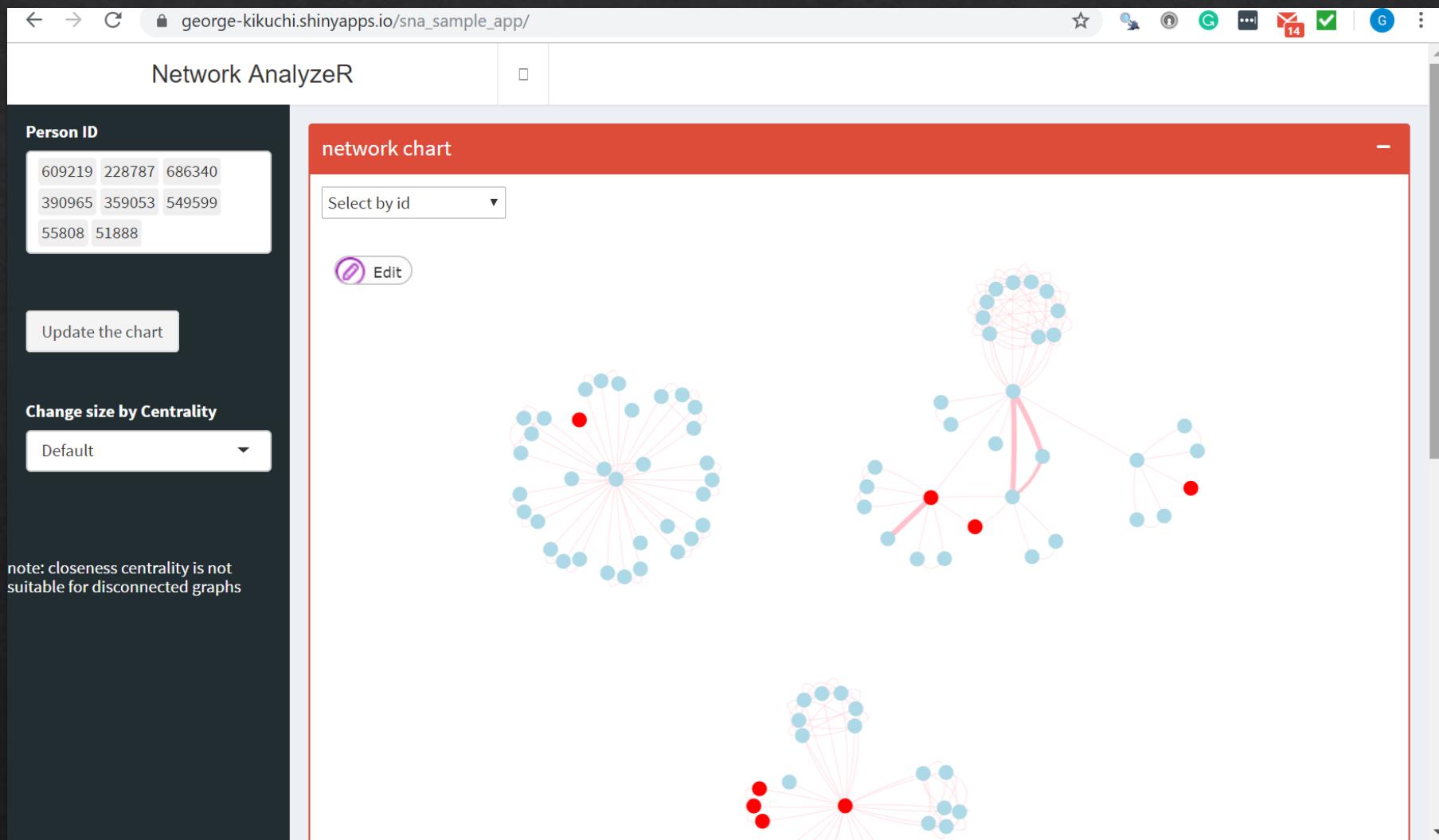




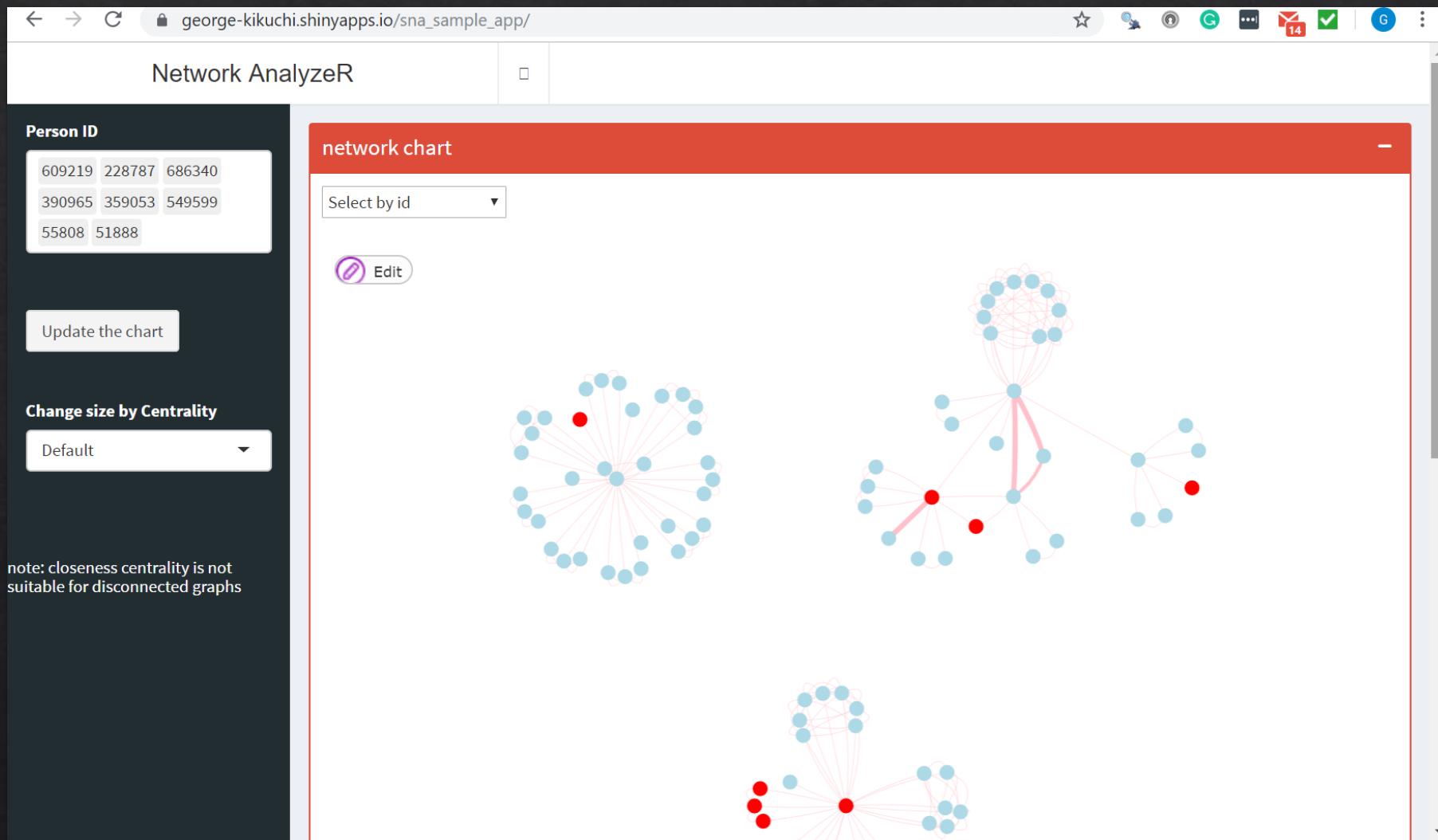
# Introduction to Social Network Analysis and SNA App

George Kikuchi ([george.kikuchi@phila.gov](mailto:george.kikuchi@phila.gov))

# Sample App



<https://bit.ly/2L6fKzh> or  
[https://george-kikuchi.shinyapps.io/sna\\_sample\\_app/](https://george-kikuchi.shinyapps.io/sna_sample_app/)



# Social Network Analysis

- ❖ A visual and statistical analysis technique to examine relationships among people (and organizations)
  - ❖ Visualize connections among criminals
  - ❖ Identify key players (and secondary targets)
  - ❖ Simulate the secondary effect of enforcement actions

# Social Network Analysis: Concepts

- ❖ Node
  - ❖ circle/square
  - ❖ People, organization, locations, etc.
- ❖ Edge
  - ❖ line
  - ❖ connections (directional/non-directional)

# Social Network Analysis: Data Sources

- ❖ Typical data sources for law enforcement are co-offending, but relationships may also come from ...
  - ❖ investigatory stops, co-habitation, kinship ties, intelligence, victim-offender, etc.
- ❖ Social Media ≠ Social Network Analysis
  - ❖ Social media can be a data source

# SNA Metrics: Centrality

- ❖ Total Degree
  - ❖ Who has the most ties?
  - ❖ Helps identify the most active offender in your network
    - ❖ Be careful about the meaning of “most active”
- ❖ Eigen-vector
  - ❖ Individual who is connected to someone who has many ties
  - ❖ “A guy who knows the guys”
- ❖ Betweenness
  - ❖ Who sits between network groups?

# SNA Metrics: Centrality

- ❖ You can use a single metric; alternatively, a more robust approach is to identify key individuals who score high across multiple metrics
- ❖ SNA metrics make sense only when your network has a single type of mode
  - ❖ e.g., all of your nodes are people
  - ❖ If your network has people connected to incidents, for example, your data or network need to be manipulated
  - ❖ The app we will build assumes that the data represent people – incidents connections in an Excel file and it internally handles this data conversion to single mode data (all people)

# Outline

- ❖ Sample Demo App
- ❖ SNA overview and concepts
- ❖ SNA app
  - ❖ Excel version
  - ❖ Access / SQL ODBC version

All free and  
open source



> 3000 libraries (packages)  
created by users for  
specific tasks (e.g., making  
a map; getting data from  
RMS (SQL))

Helps write / edit a script  
and manage projects

Core / Foundation software,  
but you nowadays do not  
directly use it

[https://github.com/george-kikuchi/iaca\\_sna\\_webinar](https://github.com/george-kikuchi/iaca_sna_webinar)

- ❖ Step-by-step tutorial (Word doc)
- ❖ This presentation file (Power Point)
- ❖ App files

The screenshot shows a GitHub repository page. At the top, the URL 'github.com/george-kikuchi/iaca\_sna\_webinar' is visible in the browser's address bar. The repository name 'george-kikuchi / iaca\_sna\_webinar' is displayed above the main content area. The repository has 4 commits, 1 branch, 0 releases, and 1 contributor. The commit history lists several files: 'george-kikuchi readme markdown', 'sna\_Access\_SQL', 'sna\_Excel', '.gitignore', 'IACA2019\_Applications\_of\_SNA.pptx', 'README.html', 'README.md', 'iaca\_sna\_webinar.Rproj', and 'tutorial\_Excel.docx'. The most recent commit is 'george-kikuchi readme markdown' from 1 hour ago. Below the commit history, there is a section titled 'README.md' with the heading 'SNA apps for IACA webinar'. The description states: 'This repository contains presentation, tutorial, and R files for an IACA webinar on Social Network Analysis. In particular, the SNA app utilizes R, RStudio and Shiny to examine criminal networks. There are two sets of apps; one starts with Excel files as an input and the other directly queries database (MS Access or SQL).'

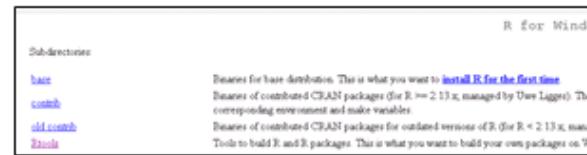
# Using the SNA app: Overall Process (detailed instructions in the Word doc on github)

- a) install R <https://cran.r-project.org/>
- b) install RStudio <https://rstudio.com/products/rstudio/download/>
- c) download [the git repository as a zip file](#)
- d) extract the downloaded zip folder
- e) go to either the sna\_Excel folder (recommended first) or  
sna\_Access\_SQL
- f) open RStudio project file (\*\*\*.Rproj)
- g) run app.R file

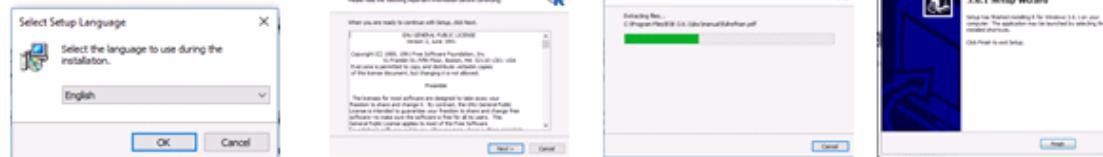
# Steps 1 – 2: Install R and RStudio

## 1. Install R (can be skipped if you already have R 3.6.1)

- Go to <https://cran.r-project.org/>, select Download R for Windows (or whatever is your OS), and click “Install R for the first time”



- Accepting all the default settings during the installation process is typically fine



## 2. Once R finishes installation, install RStudio (can be skipped if you already have RStudio)

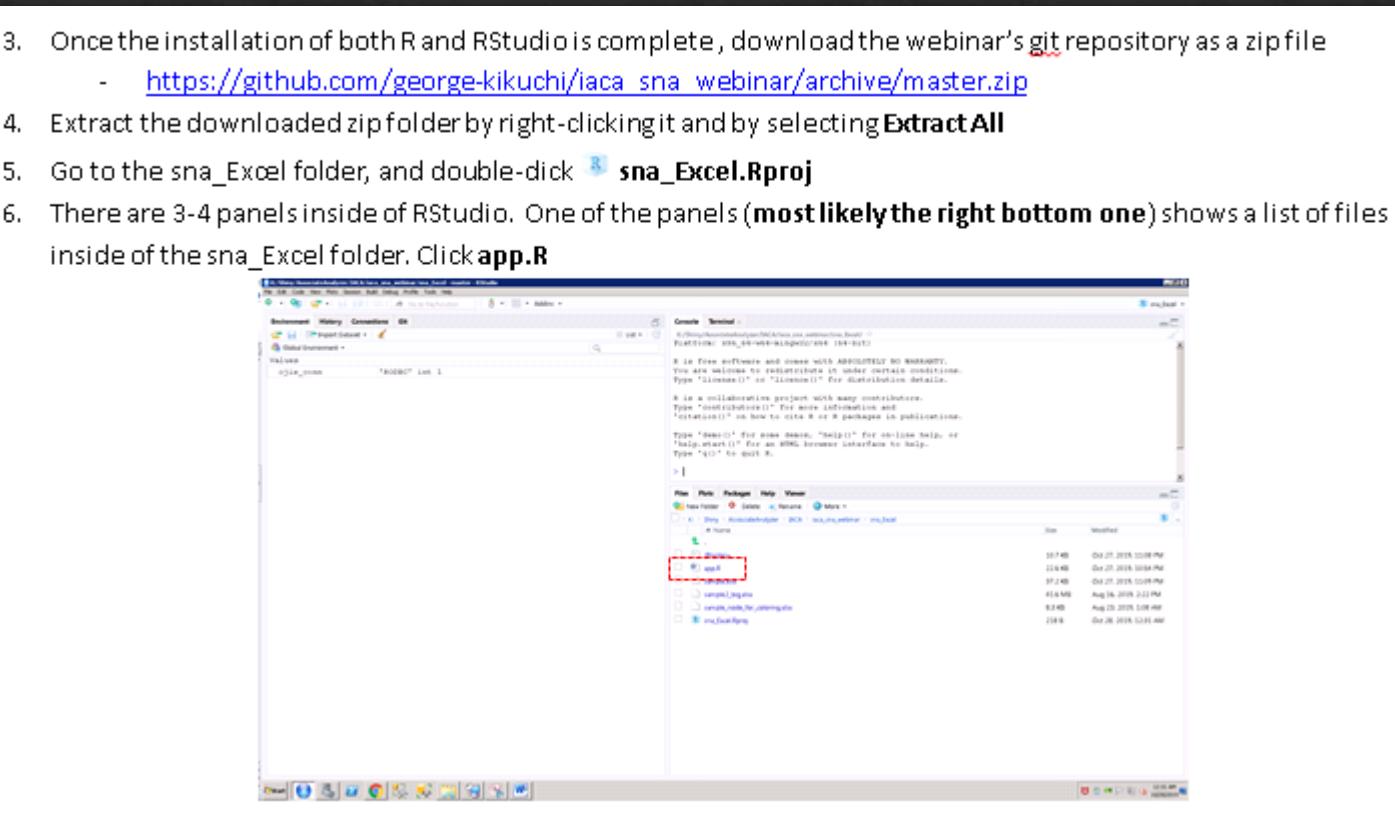
- Go to <https://rstudio.com/products/rstudio/download/#download> and click the latest version of RStudio installer for your OS (e.g., Windows 10/8/7/64-bit)
  - if you are on a 32 bit system, you will likely have to use an older version of RStudio (not tested for this app)
- Accepting all the default settings during the installation process is typically fine



# Steps 3 - 6:

## Downloading a zip folder from github, extracting it, opening sna\_Excel.Rproj file, and opening app.R

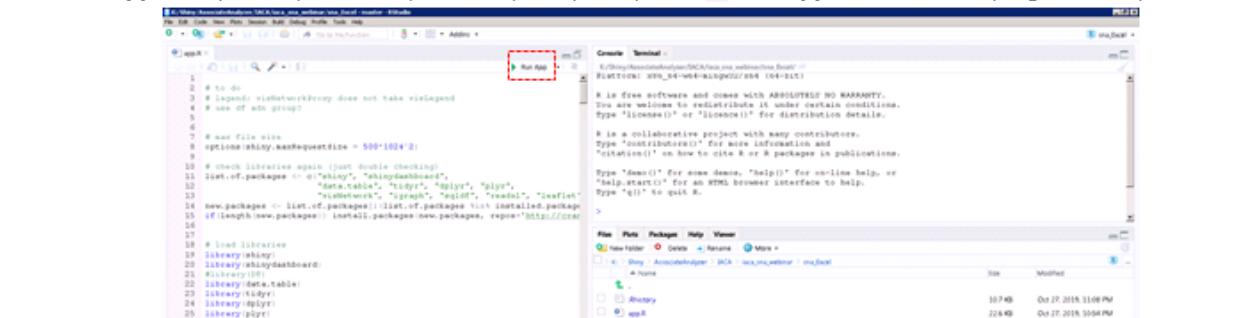
3. Once the installation of both R and RStudio is complete, download the webinar's git repository as a zip file
  - [https://github.com/george-kikuchi/iaca\\_sna\\_webinar/archive/master.zip](https://github.com/george-kikuchi/iaca_sna_webinar/archive/master.zip)
4. Extract the downloaded zip folder by right-clicking it and by selecting **Extract All**
5. Go to the **sna\_Excel** folder, and double-click  **sna\_Excel.Rproj**
6. There are 3-4 panels inside of RStudio. One of the panels (**most likely the right bottom one**) shows a list of files inside of the **sna\_Excel** folder. Click **app.R**



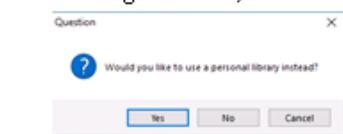
# Steps 7 - 10:

## Run the App and install necessary libraries

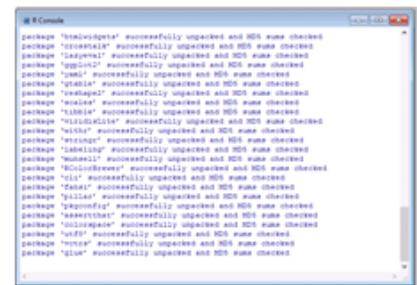
7. Once **app.R** is opened (most likely in the top left panel), click **Run App** towards the top right of this panel.



8. Only for the first time running this app, the app installs necessary libraries/packages.
  9. Depending on your right to your machine, you may be asked about using a personal library. If you see a message like this, click Yes.



10. Most of the libraries should be fairly quick to install, but depending on the speed of your internet connection, be patient with this process; these libraries just need to be installed once.



# Steps 11 - 15: Select sample Excel files

K:/Shiny/AssociateAnalyzer/IACA/iaca\_sna\_webinar/sna\_Excel - Shiny  
http://127.0.0.1:6476 | Open in Browser | ⌂

Network AnalyzeR - Excel edition

Select your Excel file that has inct\_id and person\_id

Choose Excel File for network chart

Browse... No file selected

Choose Excel File to color-code nodes (optional)

Browse... No file selected

note: closeness suitable for disc

Change size

Default

Change color

click the preview data button only after 'upload complete' is shown

Preview Data

K:/Shiny/AssociateAnalyzer/IACA/iaca\_sna\_webinar/sna\_Excel - Shiny  
http://127.0.0.1:6476 | Open in Browser | ⌂

Network AnalyzeR - Excel edition

Person ID

select column names for incident\_id, person\_id, name, and type

incident id	person id	name	type
inct_id	person_id	name2	type

Update the chart

Change size by

Default

note: closeness suitable for disc

Change color

select a person\_id column for the node attribute file (blank if no file was selected)

person id

person_id
person_id

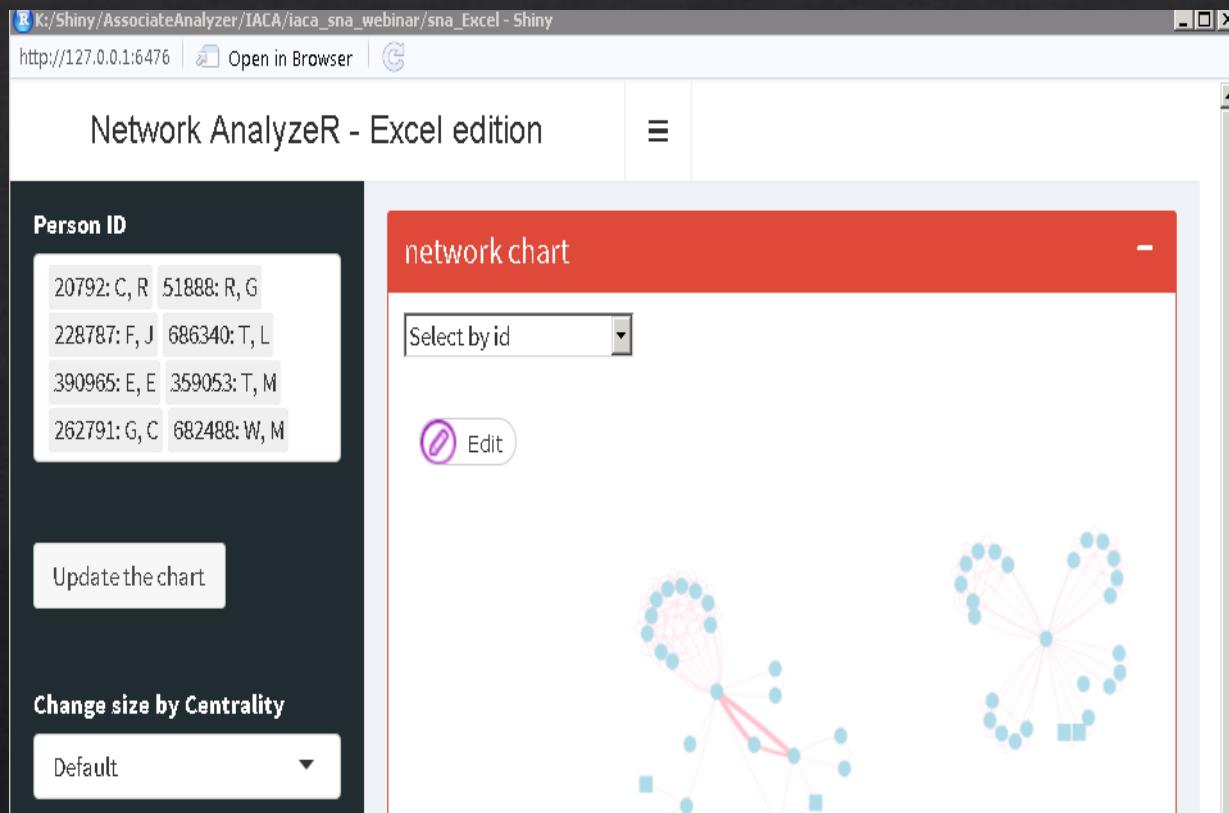
person\_id gang gun\_crime property

person_id	gang	gun_crime	property
20792.00	gang A	1.00	0.00
609219.00	gang A	1.00	0.00
51888.00	NA	3.00	0.00

Import Data

make sure column selections are not duplicated

# Steps 16: Start your analysis



# Using your own data: Excel File Format

- When using your own data (e.g., arrest record data dump from your agency's RMS), it is better to format the Excel table(s) as follows.
  - For network data (required file), your file needs to include a crime incident number (`inct_id`) and offender\_id (`person_id`) at minimum. Good column names for this file are **inct\_id**, **person\_id**, **name**, **date**, **type**. You can include more columns, but these are the required columns.

	A	B	C	D	E
1	inct_id	person_id	name	date	type
2	2000120259	20792	C, R	5/11/2000	Aggravated Assault
3	2000120259	609219	F, C	5/11/2000	Aggravated Assault
4	2005259284	51888	R, G	5/18/2005	Aggravated Assault
5	2005259284	549599	T, A	5/18/2005	Aggravated Assault

- For the optional node-coloring file (sample\_node\_for\_coloring.xlsx), the required field is **person\_id** with at least one more column that will be used for coloring. The additional column can contain either numeric or categorical values.

	A	B	C	D
1	person_id	gang	gun_crime	property
2	20792	gang A	1	0
3	609219	gang A	1	0
4	51888		3	0
5	549599		0	0
6	228787	Gang E	0	0
7	55808		0	1

# Outline

- ❖ Sample Demo App
- ❖ SNA overview and concepts
- ❖ SNA app
  - ❖ Excel version
  - ❖ Access / SQL ODBC version

# Access / SQL ODBC Connection version

Advanced version that requires....

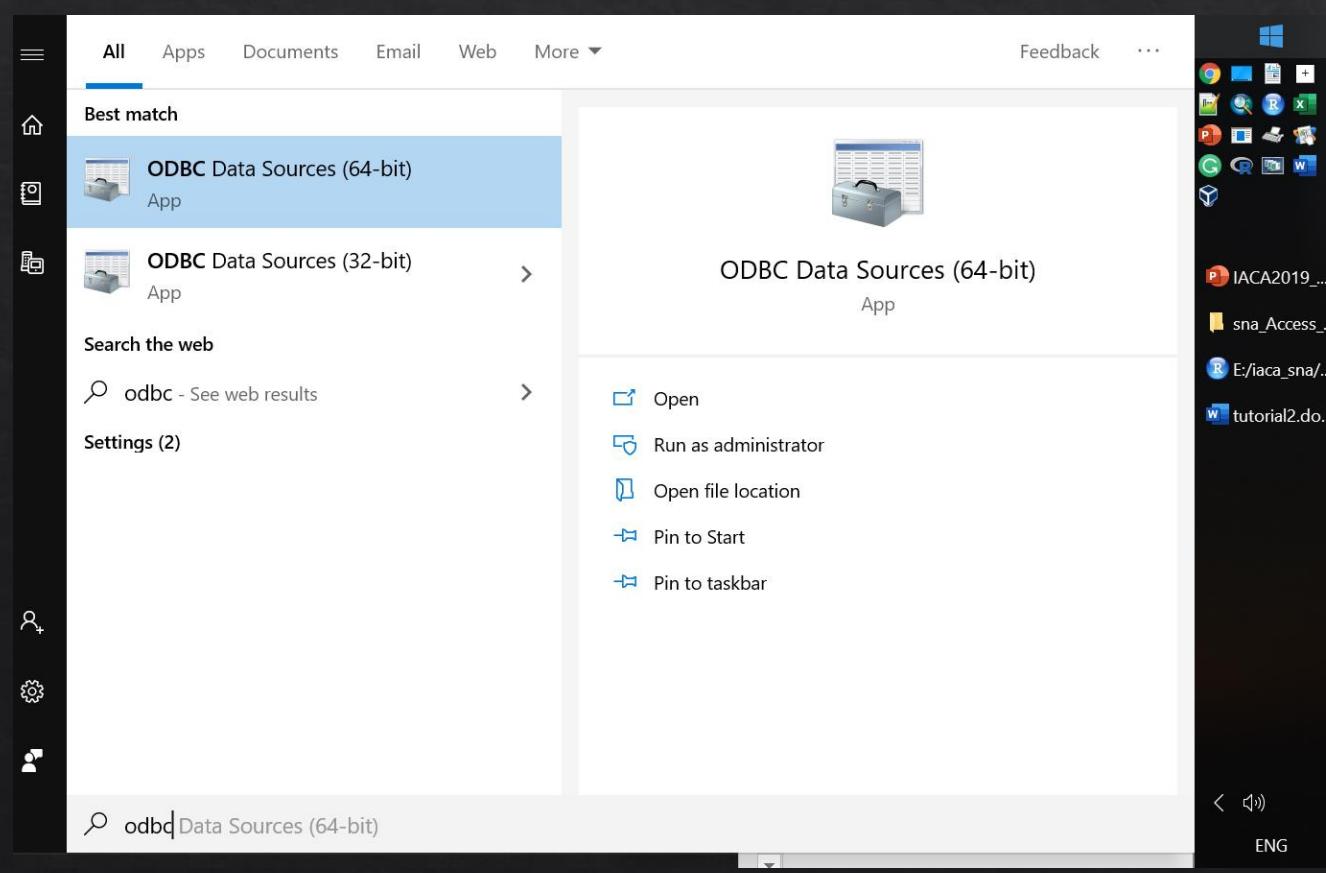
- ODBC connection string information
  - Server name (or IP address), database name, table (or view) name, Driver name (exact spelling)
- ODBC Driver installed on your computer
  - The right version for your computer and applications (R/Rstudio); 32 bit vs 64 bit
- Read access to your database server (e.g., SQL)
  - “backend access”

Advantages

- realtime and dynamic query (no periodic data dump)
- Making the app available to other people in your agency's network (behind firewall)

# Access / SQL ODBC Connection version

- ❖ Checking ODBC Driver; type ODBC in the Windows start menu and select ODBC Data Sources (XX.bit); then examine the Driver tab



The image displays two windows of the 'ODBC Data Source Administrator'. The top window is titled '(64-bit)' and the bottom window is titled '(32-bit)'. Both windows show tables of installed ODBC drivers. The columns in the tables are 'Name', 'Version', 'Company', and 'File'. In the 64-bit version, the drivers listed are:

Name	Version	Company	File
Microsoft Access dBASE Driver (*.dbf, *.ndx, *.mdx)	14.00.7180.5000	Microsoft Corporation	ACEODBC.DLL
Microsoft Access Driver (*.mdb, *.accdb)	14.00.7180.5000	Microsoft Corporation	ACEODBC.DLL

In the 32-bit version, the drivers listed are:

Name	Version	Company	File
Driver da Microsoft para arquivos texto (*.txt; *.csv)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Driver do Microsoft Access (*.mdb)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Driver do Microsoft dBase (*.dbf)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Driver do Microsoft Excel (*.xls)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Driver do Microsoft Paradox (*.db)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Microsoft Access Driver (*.mdb)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Microsoft Access-Treiber (*.mdb)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Microsoft dBase Driver (*.dbf)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL
Microsoft dBase-Treiber (*.dbf)	10.00.17763.01	Microsoft Corporation	ODBCJT32.DLL

Both windows also contain a note at the bottom: 'An ODBC driver allows ODBC-enabled programs to get information from ODBC data sources. To install new drivers, use the driver's setup program.'

# Access / SQL ODBC Connection version

- ◊ If you did not see Access driver (.accdb) under the Driver tab, you need to install necessary drivers
  - ◊ The Access driver does **not** come with Access installation (check ODBC Administrator from the windows start menu and see if Access Driver is listed under the Driver tab)
  - ◊ Try 2010 or 2007, rather than 2016 version
    - ◊<https://www.microsoft.com/en-us/download/confirmation.aspx?id=13255>
  - ◊ You also need to be mindful of 32 or 64bit version (Driver and R/Rstudio)
- ◊ Alternatively, use SQL (or equivalent)
  - ◊ Server-side processing
  - ◊ SQL Driver likely comes with Windows

# Access / SQL ODBC Connection version

- App setup (this just needs to be done once)
  - Editing ODBC connection information in **config.R** file
  - Editing the names of your table (view) and column in the **config.R** file
- iaca\_sna\_webinar → sna\_Access\_SQL →  
double-click  sna\_Access\_SQL.Rproj

# Access / SQL ODBC Connection version

The screenshot shows the RStudio interface with the following components:

- Title Bar:** E:/iaca\_sna/sna\_Access\_SQL - RStudio
- Menu Bar:** File Edit Code View Plots Session Build Debug Profile Tools Help
- Toolbar:** Standard RStudio icons for file operations.
- Code Editor:** The file `config.R*` is open. The code contains comments starting with `#`. A large orange callout bubble points to this area with the text: "2) Edit ODBC info (comments in R script start with #)".

```
1 #####
2 # configuration file
3 #####
4
5 library(DBI)
6 library(odbc)
7 library(RODBC)
8
9 #### 1) database connection string
10
11 #### sample MS access; comment these lines out with # or delete,
12 #### when using your own agency data
13 con <- DBI::dbConnect(odbc::odbc(),
14   .connection_string=paste0(
15     "Driver={Microsoft Access Driver (*.mdb, *.accdb)}",
16     "DBQ=", file.path(getwd(), "sample2_big.accdb")))
17
18 #### MS SQL sample connection;
19 #### take out # to uncomment these lines and edit server and db information
20
21 # con <- DBI::dbConnect(odbc::odbc(),
22 #   Driver = "SQL Server",
23 #   Server = "xx.xxx.xxx.xxxx",
24 #   Database = "database name",
25 #   Trusted_Connection = "True")
26
27 #### 2) table name ("edit the name within the quotes to match your data")
28 #### this table (or view) needs to have incident_id, offender_id, and name columns
29 table <- "arrest"
30
31
```
- Console:** Shows the R environment and help documentation.
- File Browser:** Shows the directory structure: E:/iaca\_sna/sna\_Access\_SQL. It includes files like `..`, `.Rhistory`, `config.R`, `sample.accdb`, `sample.laccdb`, and `sna_Access_SQL.Rproj`.

**Annotations:**

- A large orange callout bubble points to the code editor with the text: "2) Edit ODBC info (comments in R script start with #)".
- A large orange callout bubble points to the file browser with the text: "1) Open config.R".

# Access / SQL ODBC Connection version

The screenshot shows the RStudio interface with the following components:

- Top Bar:** Shows the R logo and the path "E:/iaca\_sna/sna\_Access\_SQL - RStudio".
- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New, Open, Save, Run, and Go to file/function.
- Code Editor:** Displays the "app.R" file content. The code is as follows:

```
1 options(list(scipen=999, launch.browser = TRUE)) #, shiny.port = 5155, shiny.h
2
3 # online version
4 list.of.packages <- c("shiny", "shinydashboard",
5                         "DT", "data.table",
6                         "DBI", "odbc", "RODBC",
7                         "plyr", "dplyr", "tidyverse",
8                         "visNetwork", "igraph")
9
10 new.packages <- list.of.packages[!(list.of.packages %in% installed.packages())]
11
12 # load libraries
13 library(shiny)
14 library(shinydashboard)
15 library(DT)
16 library(data.table)
17 library(tidyverse)
18 library(dplyr)
19 library(plyr)
20 library(visNetwork)
21 library(igraph)
22
23 source("config.R")
24
25 size.list <- c("Default",
26                  "betweenness", "eigen_vector",
27                  "closeness", "total_degree")
28
29
30 # Load name list
31 name.list <- dbGetQuery(con,
32                         paste0("select distinct ", person_id,
33                               " from ", table
34 ))
```

- Run App Button:** A red box highlights the "Run App" button in the toolbar.
- Help/Console Area:** Shows R's license information and a command prompt (>).
- File Browser:** Shows the directory structure: E:/iaca\_sna/sna\_Access\_SQL. It includes files ".Rhistory", "app.R", "config.R", "sample.accdb", "sample.laccdb", and "sna\_Access\_SQL.Rproj".
- Red Boxes:** Two orange boxes highlight specific areas:
  - One points to the "Run App" button with the label "2) Run App".
  - One points to the file browser with the label "1) Open app.R".



# Introduction to Social Network Analysis and SNA App

George Kikuchi ([george.kikuchi@phila.gov](mailto:george.kikuchi@phila.gov))