



# Help! I'm an Accidental Government Information Librarian

Brought to you by the North Carolina Library Association's  
Government Resources Section.

Join us! <http://www.nclaonline.org/government-resources>

# Upcoming Webinars

- February 17: Canadian statistics with Michelle Edwards, Cornell University
- March: Government apps with Emily Lawson
- April: State documents with Kris Kasianowitz
- May: [data.gov](http://data.gov) and APIs, Julia Frankosky
- June: Congressional procedure with Morgan Stoddard
  
- For more webinars, visit our website:  
<http://www.nclaonline.org/government-resources/help-im-accidental-government-information-librarian-webinars>

# MPC Data Infrastructure: Integration and Access

## Introduction to MPC data

Cathy Fitch and Katie Genadek



Minnesota Population Center

# The Minnesota Population Center

- University-wide interdisciplinary research
- Education and Training
- Revolutionary Data Tools
- Home of the IPUMS and other MPC data projects

# Minnesota Population Center

Home of the IPUMS, NHGIS, and IHIS



The MPC is one of the world's leading developers of demographic data resources. We provide population data to thousands of researchers, policymakers, teachers, and students. All MPC data are available free over the internet.

## Integrated Public Use Microdata Series



### IPUMS-International

Harmonized data for 1960 forward, covering 544 million people in 238 censuses from around the world.



### IPUMS-USA

Harmonized data on people in the U.S. census and American Community Survey, from 1850 to the present.



### IPUMS-CPS

Harmonized data on people in the Current Population Survey, every March from 1962 to the present.



### North Atlantic Population Project

Complete-count data from 1800s censuses of Canada, Great Britain, Germany, Iceland, Norway, Sweden, and the U.S.

<http://www.ipums.org/>

## Other MPC Projects



### National Historical Geographic Information System

Tabular U.S. census data and GIS boundary files from 1790 to the present.



### Integrated Health Interview Series

Annual harmonized data on people in the U.S. National Health Interview Survey from the 1960s to the present.



### American Time Use Survey-X

Annual harmonized data from 2003 forward on how U.S. adults divide their time among activities.



### Integrated Demographic and Health Series

Demographic and Health Surveys integrated for analysis across time and space, 1980s forward.



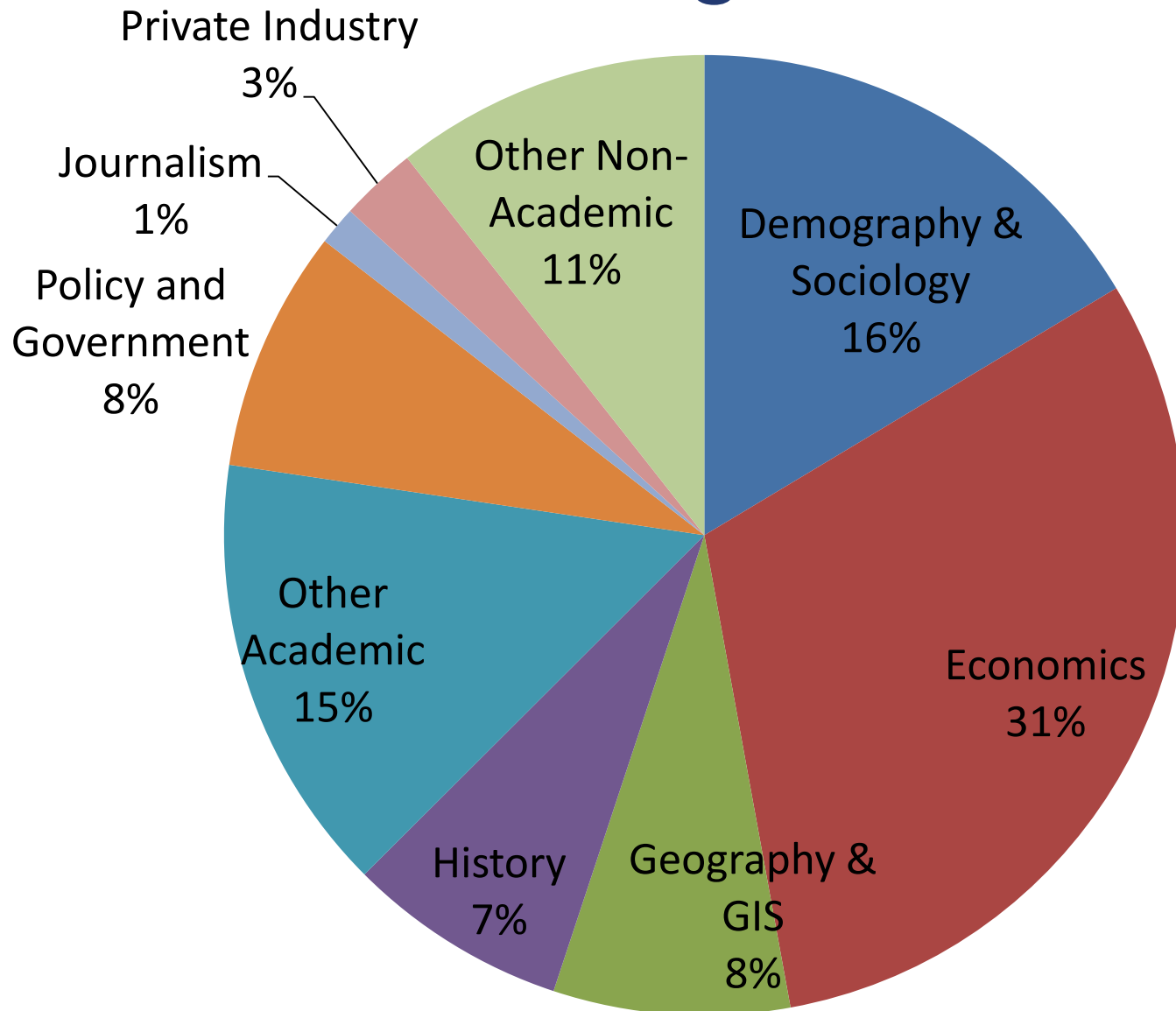
### Terra Populus

Integrated data on population and the environment, from 1960 to the present.

# What the MPC data projects do:

- Make public data freely available online
- Standardize data across time and place
- Provide “bonus” calculated fields
- Provide data in multiple formats
- Provide excellent documentation
- Provide user support

# Who is using MPC data?



# Who is using MPC data

- Scholars
- Graduate Students
- Undergraduate Students
- Local policy makers
- Researchers
- Journalists
- Government employees



# How MPC data is used:

- Explore demographic and socioeconomic change over time
- Compare demographics and other outcomes across places
- Analyze the effect of public policies on various outcomes
- Provide contextual information for other surveys and studies
- Map information using the data and provided GIS files

# What is IPUMS stand for?

**Integrated** - consistent codes, labels, and documentation

**Public Use** - anonymized, downloadable

**Microdata** - individual-level

**Series** - pooled data over time and place

# Microdata

- Individual level data
- Shows full survey responses for each respondent
- Allows for custom tables and analyses
- Geographic suppression, item-level suppression, top and bottom codes

# Microdata

Relation  
to head

Marital  
status

Literacy

Occupation

0000980001001002	1000	0791	220	220802	11	203900	999	009900000000001
0000980001002001	2000	0642	220	220802	21	203100	999	009900000000002
0000980001003000	4100	0231	100	220802	21	201000	832	00269000001202
0000980001004000	4100	0132	100	220801	21	203300	999	009900000000002
0000988001001002	1000	0351	210	220802	23	112300	723	005270000000302
0000988001002001	2000	0412	210	220802	21	203100	999	009900000000002
0000988001003000	3000	0172	100	220802	22	122300	998	009900000000002
0000988001004000	3000	0162	100	220802	22	123100	999	009900000000002
0000988001005006	4910	0192	210	220802	22	123100	999	009900000000002
0000988001006009	4910	0271	210	220802	23	121000	122	00011000009602
0000988001007000	4100	0002	100	220800	00	000000	999	00990999999992
0000988001008000	4910	0221	100	220802	23	111000	611	00011000002801
0001005001001002	1000	0451	210	220801	22	121000	611	00011000002002
0001005001002001	2000	0342	210	220801	22	13100	999	009900000000002
0001005001003000	3000	0122	100	220801	22	123300	999	009900000000002
0001013001001002	1000	0651	220	220702	21	203430	999	009900000000002
0001013001002001	2000	0652	220	220802	21	203100	999	009900000000002
0001013001003000	3000	0332	350	220802	22	13100	999	009900000000002
0001013001004000	4100	0152	100	220801	22	13300	999	009900000000002
0001013001005000	4100	0132	100	220801	22	123300	999	009900000000002
0001013001006000	4100	0021	100	220800	00	000000	999	00990999999992

Microdata Data:

**IPUMS-USA**

# IPUMS-USA

Database includes public use microdata samples:

- U.S. decennial censuses (1850-2000)
- Complete-count dataset for 1880, 1940
- Linked Samples 1850 – 1930
- Samples from Puerto Rico (1910-2014)
- American Community Survey (2000-2014)

# Upcoming USA Data Projects

1960 5% Sample

1850 full count dataset

1860 – 1930 full count individual-level datasets

1790 – 1840 full count household-level datasets

Microdata Data:

**IPUMS-CPS**



# Current Population Survey (CPS)

Administered starting 1940

Household survey was designed to measure unemployment

Source of the official Government statistics on employment and unemployment

About 57,000 households interviewed monthly

Additional Supplements given throughout year

# IPUMS - CPS

Basic Monthly Surveys (1989-2014)

All ASEC Data (1962 - 2015)

CPS Supplements:

Education

Voter Reg

Fertility

Veteran

Volunteer

Tobacco Use

Internet Use

Child Support

Food Security

Work Schedules

Job Tenure

Geographic areas identified: States, some metro areas, some counties

Microdata Data:

**IPUMS-INTERNATIONAL**

# IPUMS-International

World's largest collection of publicly available individual-level census data

Incorporates microdata from 82 countries

Samples from 1960-present

Plan to process and release data for new countries each year

Agreements with countries to process most recent round of census data

# IPUMS-I Countries Available

Argentina	5	Fiji	5	Malawi	3	Saint Lucia	2
Armenia	1	France	7	Malaysia	4	Senegal	2
Austria	4	Germany	4	Mali	2	Sierra Leone	1
Bangladesh	3	Ghana	1	Mexico	7	Slovenia	1
Belarus	1	Greece	4	Mongolia	2	South Africa	3
Bolivia	3	Guinea	2	Morocco	3	South Sudan	1
Brazil	6	Haiti	3	Nepal	1	Spain	3
Burkina Faso	3	Hungary	4	Netherlands	3	Sudan	1
Cambodia	2	India	5	Nicaragua	3	Switzerland	4
Cameroon	3	Indonesia	9	Pakistan	3	Tanzania	2
Canada	4	Iran	1	Palestine	2	Thailand	4
Chile	5	Iraq	1	Panama	6	Turkey	3
China	2	Ireland	8	Peru	2	Uganda	2
Colombia	4	Israel	3	Philippines	3	UK	2
Costa Rica	4	Italy	1	Portugal	3	USA	7
Cuba	1	Jamaica	3	Puerto Rico	5	Uruguay	5
Ecuador	6	Jordan	1	Romania	3	Venezuela	4
Egypt	2	Kenya	5	Rwanda	2	Vietnam	3
El Salvador	2	Kyrgyzstan	2				

Microdata Data:

**NAPP**

# North Atlantic Population Project

32 historical censuses from Canada, Great Britain, Iceland, Norway, Sweden, and the U.S. (1703-1951)

Complete count data for 14 censuses

When complete, will contain 130 million individual records

Linking project, across time and countries

Microdata Data:

**IHIS**



# Integrated Health Interview Series

National Health Interview Survey

Principal source of information on the health of the U.S. population

Annual survey data, 1963-2014

45,000 households sampled per year

More than 13,000 variables

Geographic identification for some major metropolitan areas and regions

Microdata Data:

**ATUS-X**

# American Time Use Survey

Funded by the BLS and fielded by the Census Bureau

Drawn from households in out-going rotation group of the Current Population Survey

Annual samples from 2003-2014

Respondents were asked about how they spent their day and reports:

- What they were doing

- Where they were

- Who was with them

Microdata:

**IDHS**

# Integrated Demographic and Health Surveys

Over 260 surveys total

From over 90 poor- & middle-income countries

Focus on women of childbearing age, their partners, and children under 5.

Covers many topics (e.g., family planning, child health and survival, domestic violence, women's status, HIV/AIDS, nutrition, and health care)

Partnership with Measure DHS, ICF International

Microdata:

**IPUMS-HIGHER ED**

# SESTAT: The Scientists and Engineers Statistical Data System

- Three surveys of U.S. residents with science or engineering degrees or occupations
- Each given every 2-3 years from 1993-2013
- Includes:
  - Demographic information
  - Education and work history
  - Employer characteristics
  - Earnings
- Fielded by National Science Foundation

# Summary Data

Age	Both sexes	Male	Female
<b>Total population</b>	<b>281,421,906</b>	<b>138,053,563</b>	<b>143,368,343</b>
Under 5 years	19,175,798	9,810,733	9,365,065
5 to 9 years	20,549,505	10,523,277	10,026,228
10 to 14 years	20,528,072	10,520,197	10,007,875
15 to 19 years	20,219,890	10,391,004	9,828,886
20 to 24 years	18,964,001	9,687,814	9,276,187
25 to 29 years	19,381,336	9,798,760	9,582,576
30 to 34 years	20,510,388	10,321,769	10,188,619
35 to 39 years	22,706,664	11,318,696	11,387,968
40 to 44 years	22,441,863	11,129,102	11,312,761
45 to 49 years	20,092,404	9,889,506	10,202,898
50 to 54 years	17,585,548	8,607,724	8,977,824
55 to 59 years	13,469,237	6,508,729	6,960,508
60 to 64 years	10,805,447	5,136,627	5,668,820
65 to 69 years	9,533,545	4,400,362	5,133,183
70 to 74 years	8,857,441	3,902,912	4,954,529
75 to 79 years	7,415,813	3,044,456	4,371,357
80 to 84 years	4,945,367	1,834,897	3,110,470
85 to 89 years	2,789,818	876,501	1,913,317
90 years and over	1,449,769	350,497	1,099,272



Summary Data:

**NHGIS**

# National Historical Geographic Information System (NHGIS)

- Summary tables for Census Data 1790-2010
- Can identify aggregate characteristics for smaller geographic areas (tract, block, etc)
- GIS Boundary files available
- Time Series tables available
- ACS data tables available

Microdata/Summary Data:

# **TERRA POP**

# TerraPop Goals

Preserve, integrate, disseminate,  
global-scale spatiotemporal data

- Population microdata
- Government land-use statistics
- Land cover data from satellite imagery
- Climate data (temperature, precipitation, cloud cover)

# Now available in TerraPop

- International Microdata for 74 countries
- Aggregate data at state/province level in each country
- GIS boundaries for 6 countries
- Raster data:
  - Global Landscapes Initiative
  - Global Land Cover 2000
  - WorldClim (1950-2000)

Obtaining Data:

# **DATA EXTRACTION**

# Online Extraction System

- Users create custom data files
  - Pick any samples of interest
  - Pick any variables of interest
- Creates custom syntax for reading the data files into SPSS, Stata, SAS, and CSV
  - Labels variables and values within the data
- Codebook available
- Record of extract is preserved on user account

# Online Extraction System

## Additional Features:

- Case Selection
- Attach Variables
- Custom Sample Size

## Learn one data set, learn how to use all MPC microdata:

- Documentation systems
- Extract systems
- Data files produced



# Online Analysis System

- High-speed tabulation software developed at UC-Berkeley
- Allows for analysis of microdata without statistical package
- All analysis performed online
- Can analyze multiple years of data
- Help guides on webpage

# Online Analysis Example

Research Question: How did the racial composition of North Carolina change between 1980-2010?

- What is the state code for North Carolina?
- How was race measured in each year?
- How can we compare race over time?

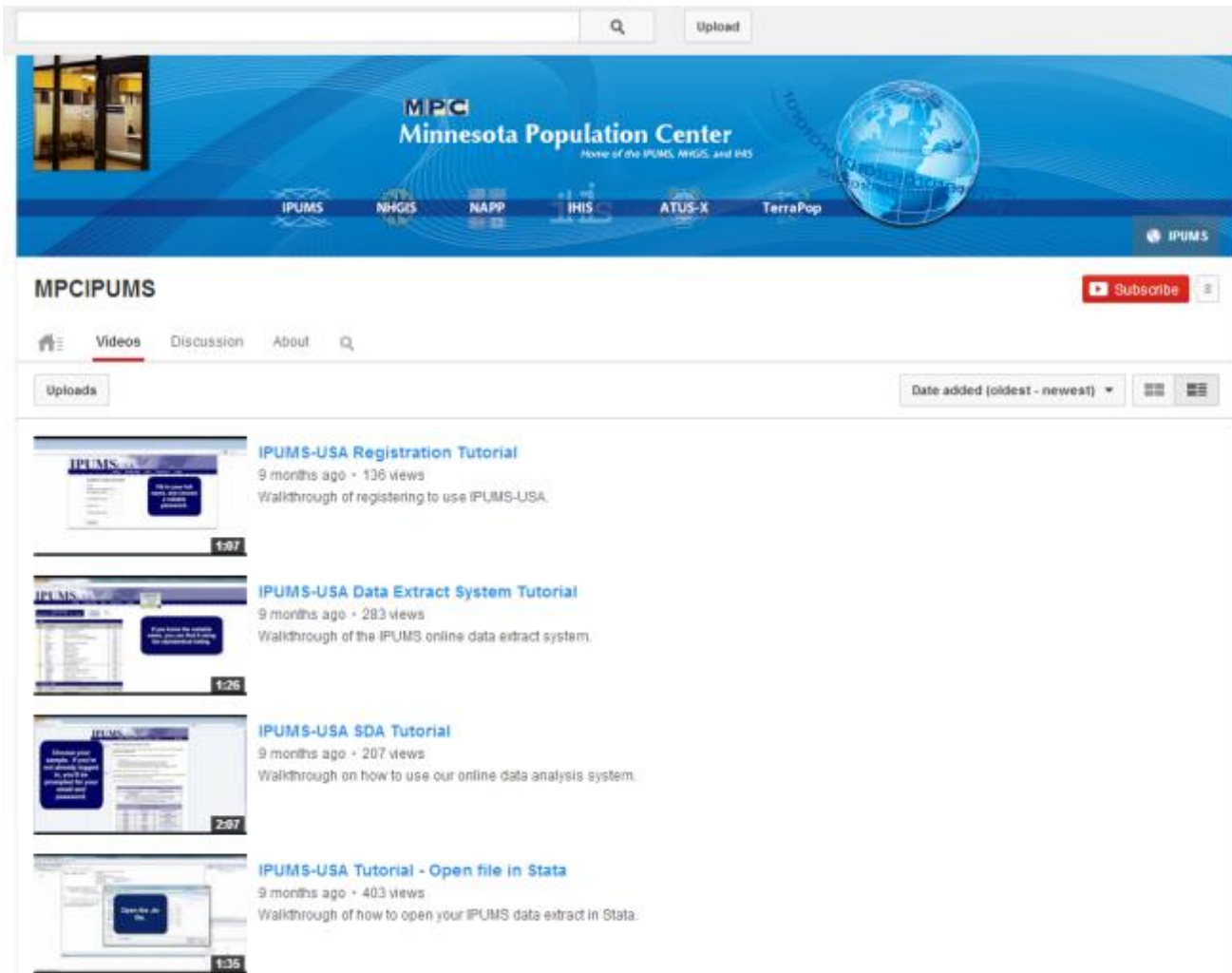
# Online Analysis Example

Research Question: In Greensboro, NC, how many people bike to work in 2014?

- Can we identify respondents in Greensboro?
- What information is there on commuting in 2014?
- Who is in the universe for the commuting question?

# Video Tutorials

<http://www.youtube.com/user/MPCIPUMS>



The screenshot displays the YouTube channel page for MPCIPUMS. At the top, there is a search bar and an 'Upload' button. The channel banner features the MPCIPUMS logo, the text 'Minnesota Population Center' (Home of the IPUMS, NHGIS, and IHS), and a globe. Below the banner, the channel name 'MPCIPUMS' is shown with a 'Subscribe' button. The 'Videos' tab is selected, showing a list of uploads. The videos listed are:

- IPUMS-USA Registration Tutorial**  
9 months ago • 136 views  
Walkthrough of registering to use IPUMS-USA.  
Duration: 1:07
- IPUMS-USA Data Extract System Tutorial**  
9 months ago • 283 views  
Walkthrough of the IPUMS online data extract system.  
Duration: 1:26
- IPUMS-USA SDA Tutorial**  
9 months ago • 207 views  
Walkthrough on how to use our online data analysis system.  
Duration: 2:07
- IPUMS-USA Tutorial - Open file in Stata**  
9 months ago • 403 views  
Walkthrough of how to open your IPUMS data extract in Stata.  
Duration: 1:36

# Training Materials

				<u>Statistical Software Package Required</u>							
Online Analysis				Exercise 1				Exercise 2			
Data Project	Online Module	Print Version 1	Print Version 2	Excel	SAS	SPSS	Stata	Excel	SAS	SPSS	Stata
IPUMS-CPS	<b>NEW</b>	X			X	X	X		X	X	X
IPUMS-USA	<b>NEW</b>	X		X	X	X	X		X	X	X
IPUMS-Int.l	<b>NEW</b>	X			X	X	X		X	X	X
IHIS	<b>NEW</b>	X			X	X	X		X	X	X
IDHS		X	X				X				X
NAPP		X			X	X	X		X	X	X
ATUS -X					X	X	X		X	X	X
NHGIS	<b>NEW</b>			X				X			
Terra Populus				X							
IPUMS-HigherEd					X	X	X		X	X	X

<https://www.pop.umn.edu/data-user-resources/data-support/>

# Minnesota Population Center

Home of the IPUMS, NHGIS, and IHIS



The MPC is one of the world's leading developers of demographic data resources. We provide population data to thousands of researchers, policymakers, teachers, and students. All MPC data are available free over the internet.

## Integrated Public Use Microdata Series



### IPUMS-International

Harmonized data for 1960 forward, covering 544 million people in 238 censuses from around the world.



### IPUMS-USA

Harmonized data on people in the U.S. census and American Community Survey, from 1850 to the present.



### IPUMS-CPS

Harmonized data on people in the Current Population Survey, every March from 1962 to the present.



### North Atlantic Population Project

Complete-count data from 1800s censuses of Canada, Great Britain, Germany, Iceland, Norway, Sweden, and the U.S.

<http://www.ipumsdata.org/>

## Other MPC Projects



### National Historical Geographic Information System

Tabular U.S. census data and GIS boundary files from 1790 to the present.



### Integrated Health Interview Series

Annual harmonized data on people in the U.S. National Health Interview Survey from the 1960s to the present.



### American Time Use Survey-X

Annual harmonized data from 2003 forward on how U.S. adults divide their time among activities.



### Integrated Demographic and Health Series

Demographic and Health Surveys integrated for analysis across time and space, 1980s forward.



### Terra Populus

Integrated data on population and the environment, from 1960 to the present.

Questions – email us:  
IPUMS User Support  
[ipums@umn.edu](mailto:ipums@umn.edu)

Cathy Fitch: [fitch@umn.edu](mailto:fitch@umn.edu)

Katie Genadek: [kgenadek@umn.edu](mailto:kgenadek@umn.edu)