Employing 3D choropleth geospatial maps to evaluate colorectal cancer incidence rates

Chelsea Zheng, Beachwood High School; Fredrick R. Schumacher, PhD, MPH, Department of Population and Quantitative Health Sciences, Case Western Reserve University

Abstract

Background

Population-based cancer registry (PBCR) publications are prominent and accurate sources of data extensively referenced by the scientific community and public health departments to monitor cancer trends across time, detect cancer clusters, and identify cancer incidence/mortality disparities. However, PBCR reports commonly employ terminology difficult for the lay public and community partners to interpret. An enduring problem for health care providers and researchers is effectively, and comprehensively, communicating cancer risk and other PBCR-related statistics to local communities with the purpose of reducing health disparities and strengthening health policies. Common data visualization tools utilized by researchers include Excel, Tablaeu, Sisense, ArcGis, R, and SAS. The recent development of several R graphical packages has harnessed this powerful language to create impactful visuals of cancer statistics widely utilized by epidemiologists and other researchers. Despite the steep learning curve and lack of a graphical user interface (GUI) in R, this statistical software package enables researchers to easily document their code and conduct reproducible research—unlike other data visualization tools with GUIs such as Excel.

Among the various cancers registered in PBCRs, colorectal cancer is one the most common causes of cancer morbidity in both men and in women. Colorectal cancer is strongly influenced by sex, with mortality rates in males significantly higher than in females. However, the ratio of colorectal cancer incidence rates between males and females differs by both tumor location and by age group; therefore larger analyses conducted with consideration of colorectal cancer and sex may be necessary to further encourage preventative measures and guide sex-specific screening protocols.

Goals

We aim to contrast sex-specific colorectal cancer incidence rates across Ohio counties employing choropleth geospatial mapping to determine statistically significant sex differences. Our secondary aim is to improve the graphical presentation of cancer incidence rates for the purpose of enhancing visualization and improve the dissemination of findings among community partners and the general public.

Materials and Methods

We extracted data from the Ohio Department of Health and the United States Cancer Statistics (USCS). The data from the USCS included age-adjusted colorectal cancer incidence rates (per 100,000 people) by county in Ohio from 2014-2018. The data from the Ohio Department of Health included the age-adjusted colorectal cancer incidence rates by Ohio county from 1996-2018, however, we restricted our analysis to only include colorectal cancer incidence rates from

2014-2018 in order to directly compare to data generated from USCS. We used R (version 4.1.0) and Rstudio (version 1.4.1717) to manipulate and visualize the data. We used the R package ‘rayshader’ (version 0.24.10) to visualize the data in 3D.

Results

The choropleth map generated within R was congruent with the choropleth map displayed by the United States Cancer Statistics’ (USCS) interactive data visualization tool. Additionally, we were able to transform these choropleth maps into 3D, movable figures with R. Of the counties where the difference between the female and male colorectal cancer incidence rate was found to be significant (p-value < 0.05), the male incidence rate was always higher than the female incidence rate, ranging from a difference of 15.4-37.3.

Conclusions

Of the 20 counties where the difference in sex-specific colorectal cancer incidence rates was statistically significant, all 20 counties had a higher male incidence rate than female incidence rate, and 17 counties were considered rural non-metropolitan statistical areas (MSAs) by the United States Office of Management and Budget. The method we used to generate 3D graphs in R could be implemented and built upon in the future to provide other researchers a convenient, reproducible, and visually impactful method of presenting data. In particular, these 3D visualizations could be incorporated into the Case Comprehensive Cancer Center website to better communicate with public partners.

Introduction

Cancer is the second leading cause of death after heart disease in the United States and significantly impacts public health (1). In order to address this issue, population-based cancer registries (PBCRs) provide statistics on the occurrence of cancer by continuously collecting and classifying information on all new cancer cases (2). PBCR publications are a prominent and accurate source of data that is extensively referenced by the scientific community to evaluate trends, patterns, and disparities. However, PBCRs may employ terminology that is often difficult for non-academia audiences to interpret. Furthermore, an enduring problem for health care providers and researchers is effectively communicating cancer risk factors to local communities with the purpose of reducing health disparities (3). Prompt, quality, and accurate health information is one of the key factors for the implementation of health policy, strengthening health coverage, and structuring health systems (4).

The ability to interpret visual patterns, trends, and correlations from visual representations of data is of increasing importance for policy-makers and researchers (2). The creation and interpretation of visual patterns is known as “knowledge translation (KT)”, which is defined by the World Health Organization (WHO) as the "exchange, synthesis, and effective communication of reliable and relevant research results” (5). Removing barriers between research and novice readers and tailoring information to target specific audiences is the core of improving KT (4). In an effort to support national policy-makers and researchers by providing a systematic and transparent method of strengthening KT, WHO launched the Evidence-informed Policy Network (EVIPNet) in 2012. EVIPNet helps translate health evidence for policy-makers by synthesizing the best available evidence in a user-friendly manner. This KT tool not only improved national health policy-making processes in all five observed countries, but EVIPNet also enabled knowledge-sharing, networking, and multi-sectoral collaborations (7).

Data visualization is a growing area of service in academic health sciences libraries and literature (8). A popular and entry-level data visualization tool is Excel, which allows users to produce simple graphs and charts (8). Other data visualization tools not well known in the healthcare field include business intelligence tools (Tablaeu, Sisense, etc.) (9) and geographical information system (GIS) software (ArcGIS, Google Earth Pro, etc.) (10). Lastly, statistical software packages such as R and SAS are popular and powerful languages that are widely used by epidemiologists and other researchers (9). Despite the steep learning curve and lack of a graphical user interface (GUI), the statistical software packages enable researchers to easily document their code and conduct reproducible research—unlike other data visualization tools with GUIs such as Excel.

Among the various cancers registered in PBCRs, colorectal cancer is one the most common causes of cancer morbidity in both men and in women. In the western hemisphere, colorectal cancer is the third most common cancer (11). Colorectal cancer can be attributed to many different risk factors, including family history, smoking, alcohol, and diet (12). Diets with red meats, preserved foods, saturated/animal fats, cholesteral, spicy foods, refined carbohydrates, have been found to have a positive association with colorectal cancer risk (13). Colorectal cancer is strongly influenced by sex, with an average male-to-female incidence rate ratio of 1.33 and male-to-female mortality ratio of 1.42 (14). However, the ratio of colorectal cancer incidence and mortality rates between males and females differ for different locations, or parts of the bowel, and for different age groups (15); therefore larger analyses conducted with consideration of colorectal cancer and sex may be necessary to further encourage preventative measures and specific screening.

Goals & Objectives

We aim to contrast the sex-specific colorectal cancer incidence rates across Ohio counties employing choropleth geospatial mapping to determine statistically significant sex differences. Our secondary aim is to improve the graphical presentation of cancer incidence rates for the purpose of enhancing visualization and improve the dissemination of findings among community partners and the general public.

Methods

We extracted data from the Ohio Department of Health (Fig. 4) and the United States Cancer Statistics (USCS) (Fig. 5 & Fig. 6). The data from the USCS included age-adjusted colorectal cancer incidence rates (per 100,000 people) in Ohio by county from 2014-2018. Given the USCS's data of cancer by sex was difficult to acquire, we only utilized the incidence and mortality rate. The data from the Ohio Department of Health included the age-adjusted colorectal cancer incidence rates by Ohio county from 1996-2018, however, we restricted our analysis to only include colorectal cancer incidence rates from 2014-2018 in order to directly compare to data generated from the USCS. We also calculated the difference between the colorectal cancer incidence rates in male and female in each county. Based on this calculation, we then applied a Wald test for a two-sided hypothesis with a p-value threshold of 0.05.

Before using the data retrieved from the Ohio Department of Health and the USCS, a few formatting adjustments were made to enable RStudio to properly import the data. Each dataset was edited so that the columns included words with no spaces and then the data was saved as tab delimited text (.txt).

The shapefile of the Ohio Department of Transportation (ODOT) county boundaries was obtained from the GEOhio Spatial Data Discovery Portal. Since this shapefile only notated each county by its corresponding ODOT county abbreviation, we utilized the ODOT county abbreviation table in order to synchronize the shapefile and the colorectal data.

We used R (version 4.1.0) and RStudio (version 1.4.1717) to manipulate and visualize the data. We used the R package ‘rayshader’ (version 0.24.10) to visualize the data in 3D. The final code that we wrote is in Figure 3.

Results

The choropleth map generated within R was congruent with the choropleth map displayed by the United States Cancer Statistics’s (USCS) interactive data visualization tool.

Additionally, we were able to transform these choropleth maps into 3D, movable figures with R (Fig. 1C).

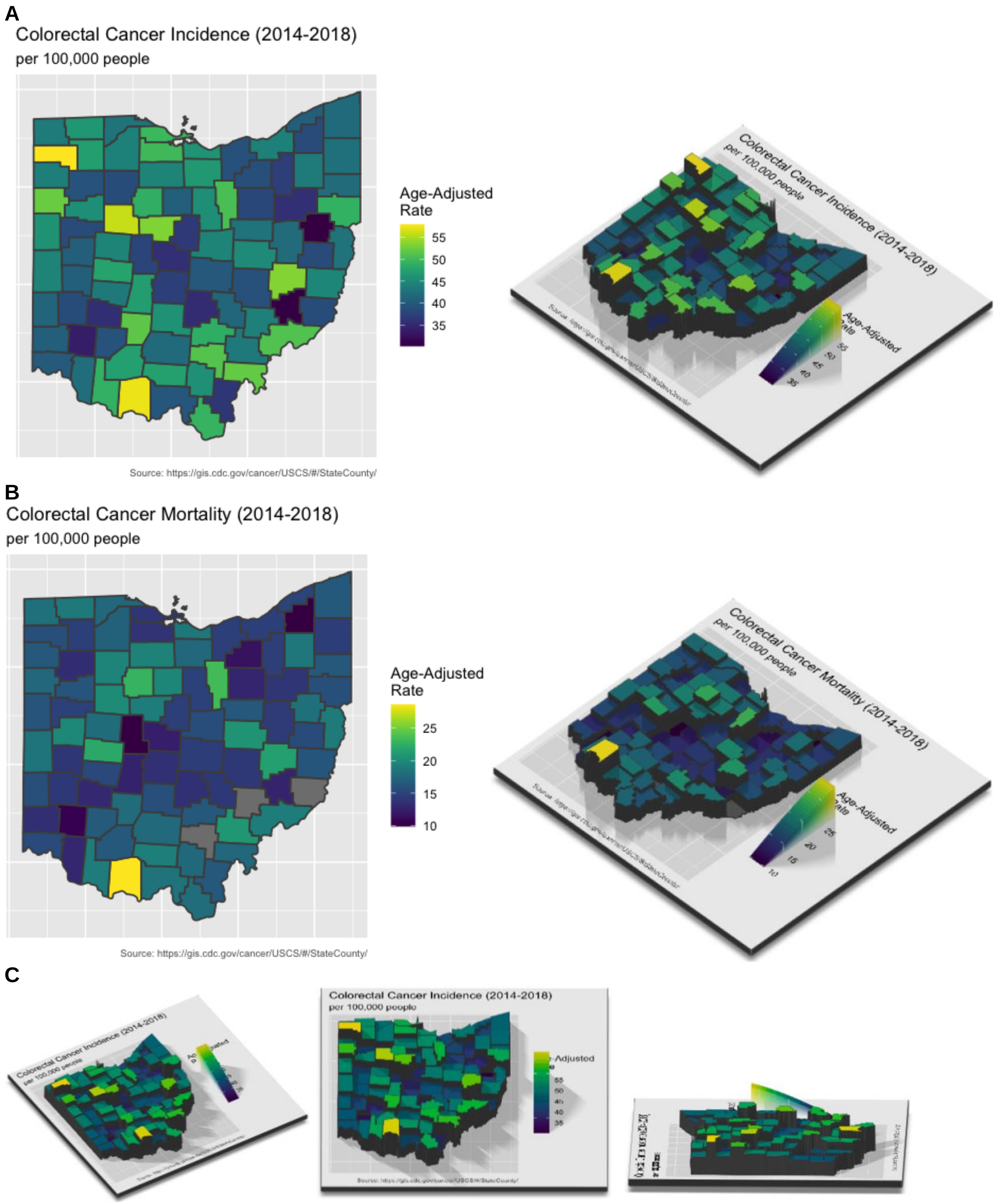

Figure 1 (A) A 2D and 3D choropleth map created in R of the age-adjusted incidence rate (per 100,000 people) of colorectal cancer in Ohio, by county. (B) A 2D and 3D choropleth map created in R of the age-adjusted mortality rate (per 100,000 people) of colorectal cancer in Ohio, by county. The grey areas indicate counties where the USCS suppressed data because fewer than 16 cases were reported. (C) A demonstration of the different perspectives that are achievable with R's 3D interactive graphs using Fig. 1A

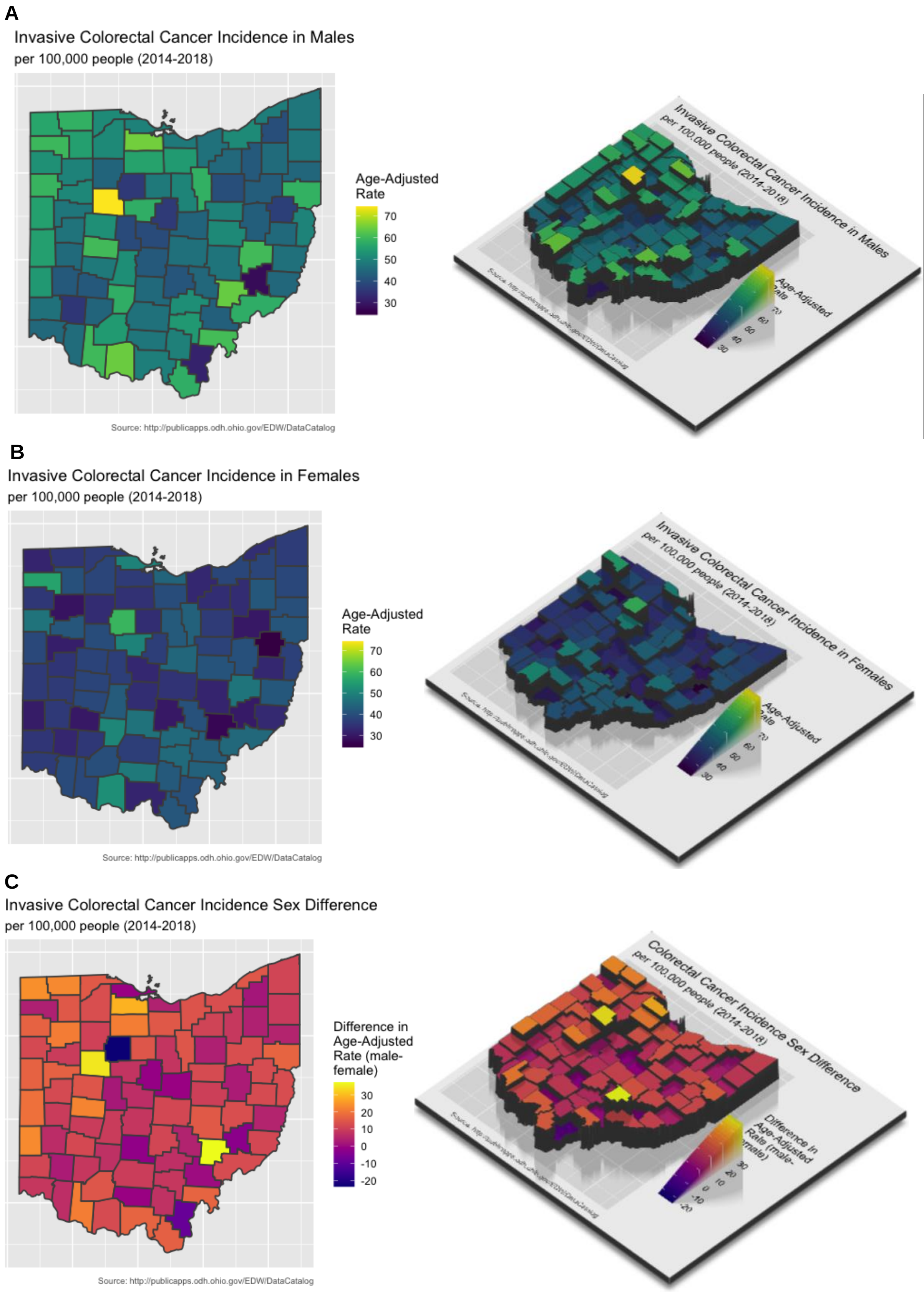
A
Invasive Colorectal Cancer Incidence in Males
per 100,000 people (2014-2018)
Age-Adjusted Rate
70
60
50
40
30
Source: http://publicapps.odh.ohio.gov/EDW/DataCatalog
Invasive Colorectal Cancer Incidence in Males
per 100,000 people (2014-2018)
B
Invasive Colorectal Cancer Incidence in Females
per 100,000 people (2014-2018)
Age-Adjusted Rate
70
60
50
40
30
Source: http://publicapps.odh.ohio.gov/EDW/DataCatalog
Invasive Colorectal Cancer Incidence in Females
per 100,000 people (2014-2018)
C
Invasive Colorectal Cancer Incidence Sex Difference
per 100,000 people (2014-2018)
Difference in Age-Adjusted Rate (male-female)
30
20
10
0
-10
-20
Source: http://publicapps.odh.ohio.gov/EDW/DataCatalog
Colorectal Cancer Incidence Sex Difference
per 100,000 people (2014-2018)

Figure 2 (A) A 2D and 3D choropleth map created in R of the age-adjusted incidence rate (per 100,000 people) of colorectal cancer in males in Ohio, by county. (B) A 2D and 3D choropleth map created in R of the age-adjusted incidence rate (per 100,000 people) of colorectal cancer in females in Ohio, by county. (C) A 2D and 3D choropleth map created in R of the difference in incidence rate of colorectal cancer between males and females in Ohio, by county (age-adjusted and per 100,000 people).

In addition to plotting the colorectal cancer incidence rates in females and males separately, we also calculated the difference between the colorectal cancer incidence rates in male and female in each county to compare the difference in incidence rate between females and males (Fig. 2C). Based on this calculation, a Wald test was applied to assess independence and a p-value for each county. Of the counties where the difference between the female and male colorectal cancer incidence rate was found to be significant (p-value < 0.05), the male incidence rate was always higher than the female incidence rate, ranging from a difference of 15.4-37.3 (Fig. 4). Of the 20 counties where the difference in sex-specific colorectal cancer incidence rates was statistically significant, 17 counties were considered rural non-metropolitan statistical areas (MSAs) by the United States Office of Management and Budget. Likewise, a majority of the counties with the highest incidence and mortality rates are also classified as rural non-MSAs.

Discussion

Colorectal cancer can be attributed to many different risk factors, including sex and environmental risk factors. In addition, choropleth maps are becoming more commonly used in the healthcare field to define the spatial location of cancer incidence and identify clusters. Our aim was to contrast the sex-specific colorectal cancer incidence rates with choropleth maps in order to examine for sex-specific differences and display these rates in a method that is more impactful then previously done.

In this study, counties with statistically significant differences could be attributed to sex-associated differences in colorectal cancer diagnosis and thus may require further attention. By comparing the choropleth maps of the incidence rates of colorectal cancer in males with the incidence rates in females, we can tell that the choropleth map of female incidence rates generally trends as a lower, darker blue —indicating a lower rate of colorectal cancer —while the male incidence rates are more elevated and display brighter greens— indicating a higher rate of colorectal cancer (Fig. 2A & Fig. 2B). The choropleth map of the incidence rate of colorectal cancer in females subtracted from the incidence rate of colorectal cancer in males is also mainly an elevated and brighter orange and yellow, indicating that the male county incidence rate was higher than the female county incidence rate (Fig. 2C). This sex disparity may be attributed to the location of the bowel that colorectal cancer is diagnosed. Patients with right-sided colorectal cancer (RCC) are more likely to be older females (16). Patients with RCC had more advanced stages at diagnoses than patients with left-sided colorectal cancer (LCC). RCC is also documented to have a lower rate of success in detection during screening (17). Additionally, the sex disparity may also be attributed to the shape of the colorectal cancer tissue. Colorectal cancer in women is more likely to appear flat and less noticeable than colorectal cancer in men (18). Last, the increased incidence rate of colorectal cancer in men may also be attributed to sex-specific anatomical characteristics in women. Women have a longer transverse colon, which may cause an incomplete colonoscopy in women (15).

The higher incidence of colorectal cancer in rural counties could be attributed to colorectal cancer screening barriers in rural areas. Barriers preventing rural residents from approaching colorectal cancer screening include lack of insurance, cost, and transportation (19). Common concerns reported by rural residents such as “fear of finding cancer”, “distrust of health care system”, and a “misperception that CRC is a male disease and CRC screening is more important for men” are psychological barriers that should be addressed by public education and awareness policies (19).

Utilizing R to create choropleth maps allowed for code to be reused to quickly create new choropleth maps with different datasets. The choropleth maps generated within R were congruent with the choropleth maps displayed by the United States Cancer Statistics’s (USCS) interactive data visualization tool (Fig. 1A & Fig. 1B). The USCS’s interactive data visualization tool only allows users to change the state, cancer, and type of data (incidence and mortality rate or number) displayed, however, by using R we were able to customize the choropleth map’s color, limits, title, etc. within R. More importantly, we were able to transform the choropleth maps into 3D interactive maps.

Implications, Community Impact, and Future Directions

Further research should be done to determine what causes sex-specific disparities in colorectal cancer incidence rate, as well as what causes the increase in colorectal incidence and mortality in rural areas. Establishing sex-specific colorectal cancer screening guidelines and breaking down barriers that prevent rural residents from colorectal cancer screening are only suggestions on what should be expanded upon in further research. In addition, the counties that were identified in this study to have statistically significant sex-specific differences may deserve further attention.

3D visualizations in R should be used as a complementary tool to communicate PBCR publications, such as cancer incidence and mortality, to policy makers and local authorities. The method we used to generate 3D graphs in R could be implemented and built upon in the future to provide other researchers a convenient, reproducible, and visually impactful method of presenting data. In particular, these 3D visualizations could be applied to other cancers and infectious diseases or incorporated into the Case Comprehensive Cancer Center website to better communicate with public partners.

Last, this study only compared the incidence and mortality rate of colorectal cancer and the sex-specific incidence rate of colorectal cancer, but further research should include sex-specific mortality rate of colorectal cancer, multiple states, and more specific or detailed regions. In the future, using an online server and the R package “Shiny” to build an interactive web app will allow people without a background in software to easily manipulate and visualize data in 3D online.

Acknowledgements

Special thanks to Dr. Schumacher for mentoring me this summer on this project and thanks to Dr. Berger, Ms. Perovsek, and all those involved with the SEO/YES program.

References

1. Centers for Disease Control and Prevention. *An Update on Cancer Deaths in the United States*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, Division of Cancer Prevention and Control; 2021.
2. Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 1857–1864. https://doi.org/10.1073/pnas.1807180116
3. King AJ, Jensen JD, Guntzviller LM, Perez Torres D, Krakow M. Ethnic newspapers and low-income Spanish-speaking adults: influence of news consumption and health motivation on cancer prevention behaviors. Ethn Health. 2018 [cited 2021 Jul 5];23(4):410-424. Available from:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6096336/#R39 doi:10.1080/13557858.2017.1280133
4. Lundkvist, A., El-Khatib, Z., Kalra, N., Pantoja, T., Leach-Kemon, K., Gapp, C., & Kuchenmüller, T. (2021). Policy-makers' views on translating burden of disease estimates in health policies: bridging the gap through data visualization. *Archives of public health = Archives belges de sante publique*, *79*(1), 17. https://doi.org/10.1186/s13690-021-00537-z
5. World report on knowledge for better health: strengthening health systems. Vol. 83. Geneva; 2005.
6. Scarlett J, Köhler K, Reinap M, Ciobanu A, Tirdea M, Koikov V, et al. Evidence-informed Policy Network (EVIPNet) Europe: success stories in knowledge translation. Public Heal Panor. 2018;04(02):161–169.
7. Ko, I., & Chang, H. (2017). Interactive Visualization of Healthcare Data Using Tableau. *Healthcare informatics research*, *23*(4), 349–354. https://doi.org/10.4258/hir.2017.23.4.349
8. LaPolla F. (2020). Excel for data visualization in academic health sciences libraries: a qualitative case study. *Journal of the Medical Library Association : JMLA*, *108*(1), 67–75. https://doi.org/10.5195/jmla.2020.749
9. Koo, H., Chun, Y., & Griffith, D. (2018). Integrating spatial data analysis functionalities in a GIS environment: Spatial Analysis using ArcGIS Engine and R (SAAR). *Transactions in GIS : TG*, *22*(3), 721–736. https://doi.org/10.1111/tgis.12452
10. Xiaopeng, Q. I., Liang, W., Barker, L., Lekiachvili, A., & Xingyou, Z. (2012). Comparison of ArcGIS and SAS Geostatistical Analyst to Estimate Population-Weighted Monthly Temperature for US Counties. *Journal of resources and ecology*, *3*(3), 220–229. https://doi.org/10.5814/j.issn.1674-764x.2012.03.004
11. Haraldsdottir, S., Einarsdottir, H. M., Smaradottir, A., Gunnlaugsson, A., & Halfdanarson, T. R. (2014). Krabbamein í ristli og endaþarmi [Colorectal cancer - review]. Laeknabladid, 100(2), 75–82. https://doi.org/10.17992/lbl.2014.02.531

12. Kuipers, E. J., Grady, W. M., Lieberman, D., Seufferlein, T., Sung, J. J., Boelens, P. G., van de Velde, C. J., & Watanabe, T. (2015). Colorectal cancer. Nature reviews. Disease primers, 1, 15065. https://doi.org/10.1038/nrdp.2015.65
13. Azeem, S., Gillani, S. W., Siddiqui, A., Jandrajupalli, S. B., Poh, V., & Syed Sulaiman, S. A. (2015). Diet and Colorectal Cancer Risk in Asia--a Systematic Review. Asian Pacific journal of cancer prevention : APJCP, 16(13), 5389–5396. https://doi.org/10.7314/apjcp.2015.16.13.5389
14. Nazarian, A., & Kulminski, A. M. (2021). Genome-Wide Analysis of Sex Disparities in the Genetic Architecture of Lung and Colorectal Cancers. Genes, 12(5), 686. https://doi.org/10.3390/genes12050686
15. Kim, S. E., Paik, H. Y., Yoon, H., Lee, J. E., Kim, N., & Sung, M. K. (2015). Sex- and gender-specific disparities in colorectal cancer risk. World journal of gastroenterology, 21(17), 5167–5175. https://doi.org/10.3748/wjg.v21.i17.5167
16. Hansen, I. O., & Jess, P. (2012). Possible better long-term survival in left versus right-sided colon cancer - a systematic review. Danish medical journal, 59(6), A4444.
17. Bressler, B., Paszat, L. F., Chen, Z., Rothwell, D. M., Vinden, C., & Rabeneck, L. (2007). Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. Gastroenterology, 132(1), 96–102. https://doi.org/10.1053/j.gastro.2006.10.027
18. Kaku, E., Oda, Y., Murakami, Y., Goto, H., Tanaka, T., Hasuda, K., Yasunaga, M., Ito, K., Sakurai, K., Fujimori, T., Hattori, M., & Sasaki, Y. (2011). Proportion of flat- and depressed-type and laterally spreading tumor among advanced colorectal neoplasia. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association, 9(6), 503–508. https://doi.org/10.1016/j.cgh.2011.03.018
19. Wang, H., Roy, S., Kim, J., Farazi, P. A., Siahpush, M., & Su, D. (2019). Barriers of colorectal cancer screening in rural USA: a systematic review. Rural and remote health, 19(3), 5181. https://doi.org/10.22605/RRH5181

Figures and Tables

Figure 3 Visualization of colorectal cancer incidence, mortality, and sex-specific differences, written in R.

```
# Read the shape file with the rgdal library.
library(rgdal)
library(rgeos)
library(mapproj)

#load Ohio map shapefile
my_spdf <- readOGR(
  dsn= "/Users/chelseazheng/Desktop/ODOT_County_Boundaries/" ,
  layer="ODOT_County_Boundaries",
  verbose=FALSE
)

# 'fortify' the data to get a dataframe format required by ggplot2
library(broom)
spdf_fortified <- tidy(my_spdf, region = "COUNTY_CD")
#the region is whatever variable in the dataset that displays the county names

# Plot empty map of Ohio to check if it worked
library(ggplot2)
```

```
basicPlot =  ggplot() +
  geom_polygon(data = spdf_fortified, aes( x = long, y = lat, group = group),
               fill="lightblue", color="white", size = 1) +
  theme_void()

basicPlot

######

#load data
Data = read.table(file = "/Users/chelseazheng/Desktop/SEO DATA/Colorectal Incidence 
USCS.txt",
                  header = TRUE, fill = TRUE)

#load ids
ODOTIDS = read.table(file = "/Users/chelseazheng/Desktop/SEO DATA/ODOT County 
Abbreviation Table.txt",
                     header = TRUE, fill = TRUE)

#Merge the data and the ids
library(dplyr)
Data = Data %>%
  left_join(. , ODOTIDS, by=c("County"="County"))

#Merge the data and the choropleth map
spdf_fortified_data = spdf_fortified %>%
  left_join(. , Data, by=c("id"="id"))

#plot final ggplot
library(viridis) #color scale
basicPlot = ggplot() +
  geom_polygon(data = spdf_fortified_data,
       aes(fill = Age_Adjusted_Rate, x = long, y = lat, group = group),
       #fill = whatever value you want to represent the color
       color="gray30", size = 0.6) +
  #get rid of x & y axis labels
  theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        plot.caption = element_text(color = "gray35", size = 7))+
  scale_fill_viridis(breaks = c(35,40,45,50,55)) +
  labs(title="Colorectal Cancer Incidence (2014-2018)", y="Latitude",
        x="Longitude", caption="Source: 
https://gis.cdc.gov/cancer/USCS/#/StateCounty/") +
  labs(subtitle="per 100,000 people")+
  labs(fill="Age-Adjusted \nRate") +
  scale_size(trans = "reverse")

basicPlot

############################################################
#### Colorectal Cancer Incidence in Females ####
Data = read.table(file = "/Users/chelseazheng/Desktop/SEO DATA/Colorectal, Invasive, 
sex.txt",
  header = TRUE, fill = TRUE)

#load ids
ODOTIDS = read.table(file = "/Users/chelseazheng/Desktop/SEO DATA/ODOT County 
Abbreviation Table.txt",
                     header = TRUE, fill = TRUE)

#Merge the data and the ids
library(dplyr)
```

```
Data = Data %>%
  left_join(. , ODOTIDS, by=c("County"="County"))

#Merge the data and the choropleth map
spdf_fortified_data = spdf_fortified %>%
  left_join(. , Data, by=c("id"="id"))

#plot final ggplot
library(viridis) #color scale
basicPlot = ggplot() +
  geom_polygon(data = spdf_fortified_data,
               aes(fill = Female, x = long, y = lat, group = group),
               #fill = whatever value you want to represent the color
               color="gray30", size = 0.6) +
  theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        plot.caption = element_text(color = "gray35", size = 7))+
  scale_fill_viridis(breaks = c(20,30,40,50,60,70,80),
                     limits = c(24.3,74.5),
                     option="viridis") +
  labs(title="Invasive Colorectal Cancer Incidence in Females", y="Latitude",
       x="Longitude", caption="Source:
http://publicapps.odh.ohio.gov/EDW/DataCatalog") +
  labs(subtitle="per 100,000 people (2014-2018)")+
  labs(fill="Age-Adjusted \nRate") +
  scale_size(trans = "reverse")

basicPlot

####differences####
basicPlot = ggplot() +
  geom_polygon(data = spdf_fortified_data,
               aes(fill = Difference, x = long, y = lat, group = group),
               #fill = whatever value you want to represent the color
               color="gray30", size = 0.6) +
  theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        plot.caption = element_text(color = "gray35", size = 7))+
  scale_fill_viridis(breaks = c(-20,-10,0,10,20,30,40),
                    #limits = c(-10,40), #wyandot?
                    option="plasma") +
  labs(title="Invasive Colorectal Cancer Incidence Sex Difference", y="Latitude",
       x="Longitude", caption="Source:
http://publicapps.odh.ohio.gov/EDW/DataCatalog") +
  labs(subtitle="per 100,000 people (2014-2018)")+
  labs(fill="Difference in \nAge-Adjusted \nRate (male- \nfemale)") +
  scale_size(trans = "reverse")
basicPlot

###########################################################
####Colorectal mortality####
Data = read.table(file = "/Users/chelseazheng/Desktop/SEO DATA/Colorectal Mortality
USCS.txt",
                  header = TRUE, fill = TRUE)

#load ids
ODOTIDS = read.table(file = "/Users/chelseazheng/Desktop/SEO DATA/ODOT County
Abbreviation Table.txt",
                     header = TRUE, fill = TRUE)

#Merge the data and the ids
```

```
library(dplyr)
Data = Data %>%
  left_join(. , ODOTIDS, by=c("County"="County"))

#Merge the data and the choropleth map
spdf_fortified_data = spdf_fortified %>%
  left_join(. , Data, by=c("id"="id"))

library(viridis) #color scale
basicPlot = ggplot() +
  geom_polygon(data = spdf_fortified_data,
               aes(fill = Age_Adjusted_Rate, x = long, y = lat, group = group),
               #fill = whatever value you want to represent the color
               color="gray30", size = 0.6) +
  #get rid of x & y labels
  theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        plot.caption = element_text(color = "gray35", size = 7))+
  scale_fill_viridis(breaks = c(5,10,15,20,25,30)) +
  labs(title="Colorectal Cancer Mortality (2014-2018)", y="Latitude",
       x="Longitude", caption="Source:
https://gis.cdc.gov/cancer/USCS/#/StateCounty/") +
  labs(subtitle="per 100,000 people")+
  labs(fill="Age-Adjusted \nRate") +
  scale_size(trans = "reverse")
basicPlot

############################################################

#turn 2d into 3d

library(rayshader)
plot_gg(basicPlot, width = 4.3, height = 4)

############################

#make the 3D plot spin
filename_movie = tempfile()
render_movie(filename = filename_movie,
             frames = 130,  phi = 45, zoom = 0.8, theta = -90)
```

Figure 4 Colorectal cancer incidence rate in males and females (age adjusted, per 100,000 people). The difference is the female incidence rate subtracted from the male incidence rate.

County	Male	Female	Total	Difference	p-value (Wald test)
Wyandot	35.7	59.3	48.9	-23.6	0.992
Gallia	31.5	40.2	36.3	-8.7	0.848
Noble	27.5	30.2	30.3	-2.7	0.639
Pickaway	45.2	46.9	45.8	-1.7	0.57
Morrow	36.0	36.9	36.5	-0.9	0.542
Knox	44.0	44.5	44.6	-0.5	0.521
Pike	47.8	48.3	47.4	-0.5	0.52
Ottawa	49.7	50.1	49.7	-0.4	0.516
Athens	43.2	42.8	43.4	0.4	0.483
Geauga	39.8	38.1	38.5	1.7	0.424
Union	39.0	36.9	37.7	2.1	0.405
Tuscarawas	43.1	40.2	41.9	2.9	0.375
Warren	35.3	32.3	33.7	3	0.358
Wayne	39.8	36.2	38.0	3.6	0.34

Defiance	59.7	56.0	57.9	3.7	0.365
Portage	45.6	41.7	43.6	3.9	0.338
Erie	48.3	44.1	46.2	4.2	0.331
Perry	44.8	40.5	41.4	4.3	0.321
Mahoning	44.6	40.1	41.9	4.5	0.312
Belmont	46.3	41.7	44.8	4.6	0.312
Montgomery	42.0	35.8	38.6	6.2	0.241
Clermont	45.7	39.0	42.2	6.7	0.233
Delaware	39.3	32.5	35.6	6.8	0.211
Hocking	52.1	45.1	48.8	7	0.239
Fayette	58.0	50.7	51.7	7.3	0.242
Jackson	49.5	42.0	45.7	7.5	0.216
Marion	58.0	50.1	53.3	7.9	0.224
Richland	50.4	42.2	45.5	8.2	0.197
Franklin	42.0	33.6	37.3	8.4	0.167
Hamilton	45.0	36.4	40.1	8.6	0.17
Vinton	54.0	45.4	51.2	8.6	0.194
Clinton	44.2	35.5	39.2	8.7	0.165
Summit	42.1	33.3	37.2	8.8	0.155
Trumbull	47.8	38.5	42.5	9.3	0.158
Allen	50.9	41.4	45.5	9.5	0.161
Madison	52.0	42.5	47.0	9.5	0.164
Miami	45.4	35.9	40.5	9.5	0.146
Clark	48.3	38.6	43.4	9.7	0.149
Ross	46.7	36.9	41.8	9.8	0.142
Total	46.9	36.6	41.3	10.3	0.13
Coshocton	51.2	40.9	45.2	10.3	0.142
Hancock	44.1	33.4	38.6	10.7	0.112
Medina	44.9	33.8	38.8	11.1	0.105
Ashtabula	47.8	36.6	41.8	11.2	0.111
Fairfield	41.3	29.9	35.2	11.4	0.088
Butler	46.7	35.1	40.5	11.6	0.1
Licking	46.4	34.7	39.9	11.7	0.097
Monroe	44.8	33.0	38.6	11.8	0.09
Greene	41.9	29.9	35.5	12	0.078
Van-Wert	59.2	47.1	52.3	12.1	0.12
Shelby	47.2	34.9	40.5	12.3	0.087
Holmes	50.1	37.7	42.7	12.4	0.093
Adams	64.5	52.0	57.3	12.5	0.123
Harrison	45.1	32.6	38.5	12.5	0.078
Carroll	37.0	24.3	30.2	12.7	0.052
Washington	57.8	45.0	51.1	12.8	0.103
Jefferson	50.2	37.2	42.8	13	0.082
Stark	42.9	29.8	35.6	13.1	0.062
Guernsey	60.4	47.3	53.4	13.1	0.103
Cuyahoga	51.4	38.2	44.2	13.2	0.082
Wood	51.3	38.1	44.3	13.2	0.081
Lorain	46.8	32.6	39.1	14.2	0.055
Logan	52.3	37.8	44.6	14.5	0.063
Ashland	58.3	43.6	50.6	14.7	0.073
Crawford	48.3	33.6	40.5	14.7	0.052
Highland	54.5	39.8	47.3	14.7	0.065
Lake	48.9	34.0	40.6	14.9	0.051
Henry	55.2	40.2	47.7	15	0.062
Huron	54.4	39.4	46.6	15	0.061

Lucas	51.5	36.1	42.8	15.4	0.05
Muskingum	47.4	32.0	39.1	15.4	0.042
Lawrence	58.0	41.4	49.3	16.6	0.048
Auglaize	47.8	31.0	38.5	16.8	0.029
Mercer	55.5	38.6	46.1	16.9	0.041
Paulding	52.4	35.5	43.9	16.9	0.036
Meigs	60.7	43.5	52.0	17.2	0.046
Scioto	49.1	31.9	39.6	17.2	0.028
Columbiana	58.5	41.2	49.1	17.3	0.042
Darke	55.3	35.9	44.6	19.4	0.021
Putnam	50.1	28.9	39.2	21.2	0.009
Brown	60.1	38.6	49.0	21.5	0.015
Seneca	55.9	34.4	44.7	21.5	0.012
Champaign	60.4	37.9	47.7	22.5	0.012
Fulton	58.3	35.2	46.1	23.1	0.008
Preble	54.0	30.5	41.3	23.5	0.005
Williams	56.3	32.1	43.6	24.2	0.005
Sandusky	65.1	36.2	50.0	28.9	0.002
Hardin	74.5	39.4	55.9	35.1	0.001
Morgan	64.2	26.9	44.1	37.3	4.63E-05

Figure 5 Colorectal cancer mortality rate (age adjusted, per 100,000 people).

Area	County	Cancer	Year	Sex	Age_Adjusted_Rate	Population
Ohio	Morgan	Colorectal	2014-2018	MaleFemale	NA	NA
Ohio	Vinton	Colorectal	2014-2018	MaleFemale	NA	NA
Ohio	Monroe	Colorectal	2014-2018	MaleFemale	NA	NA
Ohio	Geauga	Colorectal	2014-2018	MaleFemale	9.9	469544
Ohio	Union	Colorectal	2014-2018	MaleFemale	9.9	278322
Ohio	Warren	Colorectal	2014-2018	MaleFemale	10.7	1132324
Ohio	Medina	Colorectal	2014-2018	MaleFemale	11.5	886011
Ohio	Delaware	Colorectal	2014-2018	MaleFemale	12.4	985936
Ohio	Madison	Colorectal	2014-2018	MaleFemale	12.6	219926
Ohio	Noble	Colorectal	2014-2018	MaleFemale	12.7	72187
Ohio	Wayne	Colorectal	2014-2018	MaleFemale	12.8	580697
Ohio	Clermont	Colorectal	2014-2018	MaleFemale	13	1016400
Ohio	Greene	Colorectal	2014-2018	MaleFemale	13.1	828247
Ohio	Putnam	Colorectal	2014-2018	MaleFemale	13.4	169893
Ohio	Butler	Colorectal	2014-2018	MaleFemale	13.5	1890709
Ohio	Franklin	Colorectal	2014-2018	MaleFemale	13.6	6373406
Ohio	Fairfield	Colorectal	2014-2018	MaleFemale	13.6	765071
Ohio	Sandusky	Colorectal	2014-2018	MaleFemale	13.8	296452
Ohio	Ottawa	Colorectal	2014-2018	MaleFemale	13.9	203537
Ohio	Stark	Colorectal	2014-2018	MaleFemale	13.9	1866959
Ohio	Miami	Colorectal	2014-2018	MaleFemale	14	523726
Ohio	Montgomery	Colorectal	2014-2018	MaleFemale	14.1	2659150
Ohio	Perry	Colorectal	2014-2018	MaleFemale	14.1	179902
Ohio	Muskingum	Colorectal	2014-2018	MaleFemale	14.1	430346
Ohio	Tuscarawas	Colorectal	2014-2018	MaleFemale	14.4	462338
Ohio	Licking	Colorectal	2014-2018	MaleFemale	14.5	861297
Ohio	Trumbull	Colorectal	2014-2018	MaleFemale	14.5	1008836
Ohio	Summit	Colorectal	2014-2018	MaleFemale	14.6	2708103
Ohio	Henry	Colorectal	2014-2018	MaleFemale	14.7	136630
Ohio	Cuyahoga	Colorectal	2014-2018	MaleFemale	14.7	6265383
Ohio	Morrow	Colorectal	2014-2018	MaleFemale	14.7	174826
Ohio	Lorain	Colorectal	2014-2018	MaleFemale	14.7	1532981
Ohio	Auglaize	Colorectal	2014-2018	MaleFemale	14.8	228726
Ohio	Richland	Colorectal	2014-2018	MaleFemale	14.8	606346
Ohio	Lake	Colorectal	2014-2018	MaleFemale	15	1149479
Ohio	Clark	Colorectal	2014-2018	MaleFemale	15.3	675857
Ohio	Hamilton	Colorectal	2014-2018	MaleFemale	15.4	4058091
Ohio	Harrison	Colorectal	2014-2018	MaleFemale	15.5	76535
Ohio	Knox	Colorectal	2014-2018	MaleFemale	15.6	306081

Ohio	Defiance	Colorectal	2014-2018	MaleFemale	15.8	191236
Ohio	Columbiana	Colorectal	2014-2018	MaleFemale	15.8	519687
Ohio	Preble	Colorectal	2014-2018	MaleFemale	15.8	206076
Ohio	Pickaway	Colorectal	2014-2018	MaleFemale	15.8	287055
Ohio	Allen	Colorectal	2014-2018	MaleFemale	16.1	518374
Ohio	Mercer	Colorectal	2014-2018	MaleFemale	16.1	203978
Ohio	Belmont	Colorectal	2014-2018	MaleFemale	16.2	342481
Ohio	Gallia	Colorectal	2014-2018	MaleFemale	16.4	151005
Ohio	Hocking	Colorectal	2014-2018	MaleFemale	16.4	142429
Ohio	Highland	Colorectal	2014-2018	MaleFemale	16.4	215014
Ohio	Holmes	Colorectal	2014-2018	MaleFemale	16.4	219331
Ohio	Mahoning	Colorectal	2014-2018	MaleFemale	16.4	1154513
Ohio	Paulding	Colorectal	2014-2018	MaleFemale	16.4	94346
Ohio	Ashtabula	Colorectal	2014-2018	MaleFemale	16.5	490937
Ohio	Clinton	Colorectal	2014-2018	MaleFemale	16.6	209620
Ohio	Wood	Colorectal	2014-2018	MaleFemale	17.1	649268
Ohio	Meigs	Colorectal	2014-2018	MaleFemale	17.5	115758
Ohio	Huron	Colorectal	2014-2018	MaleFemale	17.5	292041
Ohio	Van-Wert	Colorectal	2014-2018	MaleFemale	17.5	141366
Ohio	Lucas	Colorectal	2014-2018	MaleFemale	17.5	2161748
Ohio	Jackson	Colorectal	2014-2018	MaleFemale	17.6	162600
Ohio	Ross	Colorectal	2014-2018	MaleFemale	17.7	385171
Ohio	Fayette	Colorectal	2014-2018	MaleFemale	17.8	143223
Ohio	Logan	Colorectal	2014-2018	MaleFemale	17.9	226381
Ohio	Lawrence	Colorectal	2014-2018	MaleFemale	18	302995
Ohio	Carroll	Colorectal	2014-2018	MaleFemale	18.1	137874
Ohio	Williams	Colorectal	2014-2018	MaleFemale	18.2	184585
Ohio	Scioto	Colorectal	2014-2018	MaleFemale	18.3	382360
Ohio	Darke	Colorectal	2014-2018	MaleFemale	18.5	258637
Ohio	Portage	Colorectal	2014-2018	MaleFemale	18.6	812553
Ohio	Erie	Colorectal	2014-2018	MaleFemale	18.7	375441
Ohio	Fulton	Colorectal	2014-2018	MaleFemale	18.8	211542
Ohio	Jefferson	Colorectal	2014-2018	MaleFemale	18.9	334407
Ohio	Brown	Colorectal	2014-2018	MaleFemale	19	218361
Ohio	Marion	Colorectal	2014-2018	MaleFemale	19.1	327192
Ohio	Washington	Colorectal	2014-2018	MaleFemale	19.1	303380
Ohio	Pike	Colorectal	2014-2018	MaleFemale	19.2	140411
Ohio	Seneca	Colorectal	2014-2018	MaleFemale	19.5	277381
Ohio	Hancock	Colorectal	2014-2018	MaleFemale	19.6	378683
Ohio	Crawford	Colorectal	2014-2018	MaleFemale	19.7	210020
Ohio	Shelby	Colorectal	2014-2018	MaleFemale	19.9	244150
Ohio	Hardin	Colorectal	2014-2018	MaleFemale	20.1	157545
Ohio	Coshocton	Colorectal	2014-2018	MaleFemale	20.7	182857
Ohio	Athens	Colorectal	2014-2018	MaleFemale	20.8	329043
Ohio	Guernsey	Colorectal	2014-2018	MaleFemale	21.4	196370
Ohio	Champaign	Colorectal	2014-2018	MaleFemale	22.1	194418
Ohio	Wyandot	Colorectal	2014-2018	MaleFemale	22.8	110511
Ohio	Ashland	Colorectal	2014-2018	MaleFemale	23.4	267359
Ohio	Adams	Colorectal	2014-2018	MaleFemale	28.6	139246

Figure 6 Colorectal cancer incidence rate (age adjusted, per 100,000 people).

Area	County	Cancer_Type	Year	Sex	Age_Adjusted_Rate	Population
Ohio	Carroll	Colon&Rectum	2014-2018	Male&Female	30.2	137874
Ohio	Noble	Colon&Rectum	2014-2018	Male&Female	30.3	72187
Ohio	Warren	Colon&Rectum	2014-2018	Male&Female	33.7	1132324
Ohio	Fairfield	Colon&Rectum	2014-2018	Male&Female	35.2	765071
Ohio	Greene	Colon&Rectum	2014-2018	Male&Female	35.5	828247
Ohio	Stark	Colon&Rectum	2014-2018	Male&Female	35.6	1866959
Ohio	Delaware	Colon&Rectum	2014-2018	Male&Female	35.6	985936
Ohio	Gallia	Colon&Rectum	2014-2018	Male&Female	36.3	151005
Ohio	Morrow	Colon&Rectum	2014-2018	Male&Female	36.5	174826
Ohio	Summit	Colon&Rectum	2014-2018	Male&Female	37.2	2708103
Ohio	Franklin	Colon&Rectum	2014-2018	Male&Female	37.3	6373406
Ohio	Union	Colon&Rectum	2014-2018	Male&Female	37.7	278322
Ohio	Wayne	Colon&Rectum	2014-2018	Male&Female	38	580697
Ohio	Auglaize	Colon&Rectum	2014-2018	Male&Female	38.5	228726
Ohio	Harrison	Colon&Rectum	2014-2018	Male&Female	38.5	76535
Ohio	Geauga	Colon&Rectum	2014-2018	Male&Female	38.5	469544

Ohio	Hancock	Colon&Rectum	2014-2018	Male&Female	38.6	378683
Ohio	Monroe	Colon&Rectum	2014-2018	Male&Female	38.6	70411
Ohio	Montgomery	Colon&Rectum	2014-2018	Male&Female	38.6	2659150
Ohio	Medina	Colon&Rectum	2014-2018	Male&Female	38.8	886011
Ohio	Lorain	Colon&Rectum	2014-2018	Male&Female	39.1	1532981
Ohio	Muskingum	Colon&Rectum	2014-2018	Male&Female	39.1	430346
Ohio	Clinton	Colon&Rectum	2014-2018	Male&Female	39.2	209620
Ohio	Putnam	Colon&Rectum	2014-2018	Male&Female	39.2	169893
Ohio	Scioto	Colon&Rectum	2014-2018	Male&Female	39.6	382360
Ohio	Licking	Colon&Rectum	2014-2018	Male&Female	39.9	861297
Ohio	Hamilton	Colon&Rectum	2014-2018	Male&Female	40.1	4058091
Ohio	Butler	Colon&Rectum	2014-2018	Male&Female	40.5	1890709
Ohio	Miami	Colon&Rectum	2014-2018	Male&Female	40.5	523726
Ohio	Shelby	Colon&Rectum	2014-2018	Male&Female	40.5	244150
Ohio	Crawford	Colon&Rectum	2014-2018	Male&Female	40.5	210020
Ohio	Lake	Colon&Rectum	2014-2018	Male&Female	40.6	1149479
Ohio	Preble	Colon&Rectum	2014-2018	Male&Female	41.3	206076
Ohio	Perry	Colon&Rectum	2014-2018	Male&Female	41.4	179902
Ohio	Ashtabula	Colon&Rectum	2014-2018	Male&Female	41.8	490937
Ohio	Ross	Colon&Rectum	2014-2018	Male&Female	41.8	385171
Ohio	Mahoning	Colon&Rectum	2014-2018	Male&Female	41.9	1154513
Ohio	Tuscarawas	Colon&Rectum	2014-2018	Male&Female	41.9	462338
Ohio	Clermont	Colon&Rectum	2014-2018	Male&Female	42.2	1016400
Ohio	Trumbull	Colon&Rectum	2014-2018	Male&Female	42.5	1008836
Ohio	Holmes	Colon&Rectum	2014-2018	Male&Female	42.7	219331
Ohio	Jefferson	Colon&Rectum	2014-2018	Male&Female	42.8	334407
Ohio	Lucas	Colon&Rectum	2014-2018	Male&Female	42.8	2161748
Ohio	Athens	Colon&Rectum	2014-2018	Male&Female	43.4	329043
Ohio	Clark	Colon&Rectum	2014-2018	Male&Female	43.4	675857
Ohio	Portage	Colon&Rectum	2014-2018	Male&Female	43.6	812553
Ohio	Williams	Colon&Rectum	2014-2018	Male&Female	43.6	184585
Ohio	Paulding	Colon&Rectum	2014-2018	Male&Female	43.9	94346
Ohio	Morgan	Colon&Rectum	2014-2018	Male&Female	44.1	73423
Ohio	Cuyahoga	Colon&Rectum	2014-2018	Male&Female	44.2	6265383
Ohio	Wood	Colon&Rectum	2014-2018	Male&Female	44.3	649268
Ohio	Logan	Colon&Rectum	2014-2018	Male&Female	44.6	226381
Ohio	Knox	Colon&Rectum	2014-2018	Male&Female	44.6	306081
Ohio	Darke	Colon&Rectum	2014-2018	Male&Female	44.6	258637
Ohio	Seneca	Colon&Rectum	2014-2018	Male&Female	44.7	277381
Ohio	Belmont	Colon&Rectum	2014-2018	Male&Female	44.8	342481
Ohio	Coshocton	Colon&Rectum	2014-2018	Male&Female	45.2	182857
Ohio	Allen	Colon&Rectum	2014-2018	Male&Female	45.5	518374
Ohio	Richland	Colon&Rectum	2014-2018	Male&Female	45.5	606346
Ohio	Jackson	Colon&Rectum	2014-2018	Male&Female	45.7	162600
Ohio	Pickaway	Colon&Rectum	2014-2018	Male&Female	45.8	287055
Ohio	Mercer	Colon&Rectum	2014-2018	Male&Female	46.1	203978
Ohio	Erie	Colon&Rectum	2014-2018	Male&Female	46.2	375441
Ohio	Fulton	Colon&Rectum	2014-2018	Male&Female	46.4	211542
Ohio	Huron	Colon&Rectum	2014-2018	Male&Female	46.6	292041
Ohio	Madison	Colon&Rectum	2014-2018	Male&Female	47	219926
Ohio	Henry	Colon&Rectum	2014-2018	Male&Female	47.2	136630
Ohio	Highland	Colon&Rectum	2014-2018	Male&Female	47.3	215014
Ohio	Pike	Colon&Rectum	2014-2018	Male&Female	47.4	140411
Ohio	Champaign	Colon&Rectum	2014-2018	Male&Female	47.7	194418
Ohio	Hocking	Colon&Rectum	2014-2018	Male&Female	48.8	142429
Ohio	Wyandot	Colon&Rectum	2014-2018	Male&Female	48.9	110511
Ohio	Brown	Colon&Rectum	2014-2018	Male&Female	49	218361
Ohio	Columbiana	Colon&Rectum	2014-2018	Male&Female	49.1	519687
Ohio	Lawrence	Colon&Rectum	2014-2018	Male&Female	49.3	302995
Ohio	Ottawa	Colon&Rectum	2014-2018	Male&Female	49.7	203537
Ohio	Sandusky	Colon&Rectum	2014-2018	Male&Female	50	296452
Ohio	Ashland	Colon&Rectum	2014-2018	Male&Female	50.6	267359
Ohio	Washington	Colon&Rectum	2014-2018	Male&Female	51.1	303380
Ohio	Vinton	Colon&Rectum	2014-2018	Male&Female	51.2	65541
Ohio	Fayette	Colon&Rectum	2014-2018	Male&Female	51.7	143223
Ohio	Meigs	Colon&Rectum	2014-2018	Male&Female	52	115758
Ohio	Van-Wert	Colon&Rectum	2014-2018	Male&Female	52.3	141366
Ohio	Marion	Colon&Rectum	2014-2018	Male&Female	53.3	327192
Ohio	Guernsey	Colon&Rectum	2014-2018	Male&Female	53.4	196370
Ohio	Hardin	Colon&Rectum	2014-2018	Male&Female	55.9	157545

Ohio	Adams	Colon&Rectum	2014-2018	Male&Female	57.3	139246
Ohio	Defiance	Colon&Rectum	2014-2018	Male&Female	57.9	191236