Response and adverse events to chemotherapy: A mock study

Table of Contents

to do: what is fu\_time when fu\_stat is 1?

# confirm only boston site data  
mockdata %>%   
 distinct(site)

## # A tibble: 18 x 1  
## site   
## <chr>   
## 1 Portland   
## 2 Ann Arbor   
## 3 St. Louis   
## 4 Rochester   
## 5 New Haven   
## 6 Boston   
## 7 Chapel Hill   
## 8 Gainesville   
## 9 Denver   
## 10 Los Angeles   
## 11 Seattle   
## 12 New York   
## 13 Madrid   
## 14 Barcelona   
## 15 Rio de Janeiro  
## 16 Sao Paulo   
## 17 Mexico City   
## 18 Nur-Sultan

# smell test (n's, NAs, etc)  
skimr::skim(mockdata) %>%   
 skimr::kable()

Skim summary statistics  
n obs: 1499  
n variables: 26

Variable type: character

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| variable | missing | complete | n | min | max | empty | n\_unique |
| age\_ord | 0 | 1499 | 1499 | 5 | 5 | 0 | 8 |
| arm | 0 | 1499 | 1499 | 6 | 9 | 0 | 3 |
| country | 0 | 1499 | 1499 | 3 | 10 | 0 | 5 |
| ethnicity | 0 | 1499 | 1499 | 8 | 33 | 0 | 6 |
| first\_name | 0 | 1499 | 1499 | 2 | 11 | 0 | 663 |
| last\_name | 0 | 1499 | 1499 | 1 | 19 | 0 | 1256 |
| name | 0 | 1499 | 1499 | 7 | 28 | 0 | 1495 |
| race | 0 | 1499 | 1499 | 5 | 16 | 0 | 7 |
| sex | 0 | 1499 | 1499 | 4 | 6 | 0 | 2 |
| site | 0 | 1499 | 1499 | 6 | 14 | 0 | 18 |

Variable type: factor

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| variable | missing | complete | n | n\_unique | top\_counts | ordered |
| fu\_fct | 0 | 1499 | 1499 | 2 | Die: 1356, Liv: 143, NA: 0 | FALSE |

Variable type: numeric

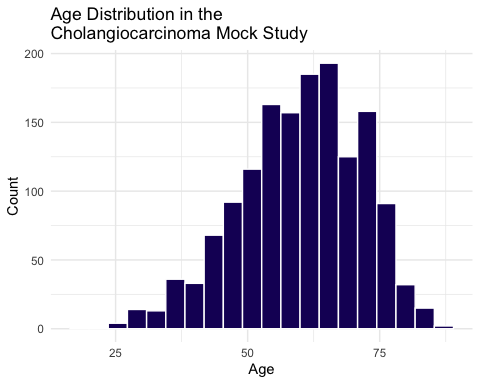
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | missing | complete | n | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| ae\_blood\_clot | 0 | 1499 | 1499 | 0.051 | 0.22 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▁ |
| ae\_diarrhea | 0 | 1499 | 1499 | 0.2 | 0.4 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▂ |
| ae\_low\_wbc | 0 | 1499 | 1499 | 0.17 | 0.38 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▂ |
| ae\_neuropathy | 0 | 1499 | 1499 | 0.15 | 0.36 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▂ |
| ae\_vomiting | 0 | 1499 | 1499 | 0.18 | 0.39 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▂ |
| age | 0 | 1499 | 1499 | 59.99 | 11.52 | 19 | 52 | 61 | 68 | 88 | ▁▁▂▅▇▇▅▁ |
| alk\_phos | 266 | 1233 | 1499 | 168.97 | 128.49 | 7 | 86 | 123 | 207 | 1014 | ▇▅▂▁▁▁▁▁ |
| ast | 266 | 1233 | 1499 | 35.93 | 26.84 | 5 | 20 | 27 | 41 | 205 | ▇▃▁▁▁▁▁▁ |
| bmi | 33 | 1466 | 1499 | 27.21 | 5.43 | 14.05 | 23.54 | 26.33 | 30.15 | 60.24 | ▁▇▇▃▁▁▁▁ |
| case | 0 | 1499 | 1499 | 98225.65 | 9256.93 | 76170 | 91305.5 | 94773 | 106859.5 | 112488 | ▁▁▂▇▂▂▃▆ |
| fu\_stat | 0 | 1499 | 1499 | 1.9 | 0.29 | 1 | 2 | 2 | 2 | 2 | ▁▁▁▁▁▁▁▇ |
| fu\_time | 0 | 1499 | 1499 | 649.08 | 462.51 | 0 | 309.5 | 542 | 878.5 | 2472 | ▆▇▅▂▂▁▁▁ |
| hgb | 266 | 1233 | 1499 | 12.35 | 1.72 | 9 | 11.1 | 12.3 | 13.5 | 18.2 | ▃▆▇▇▅▂▁▁ |
| mdquality\_s | 252 | 1247 | 1499 | 0.9 | 0.3 | 0 | 1 | 1 | 1 | 1 | ▁▁▁▁▁▁▁▇ |
| ps | 266 | 1233 | 1499 | 0.54 | 0.6 | 0 | 0 | 0 | 1 | 2 | ▇▁▁▇▁▁▁▁ |

# check for duplicates  
# hooray  
mockdata %>%   
 janitor::get\_dupes(case)

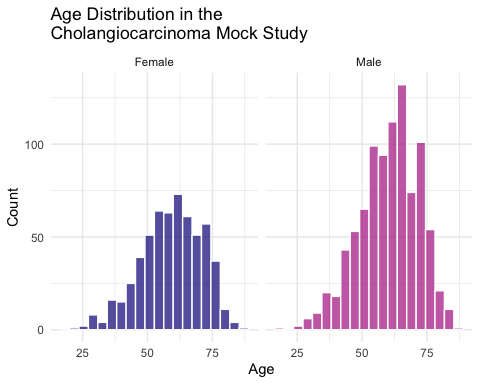
## # A tibble: 0 x 27  
## # … with 27 variables: case <dbl>, dupe\_count <int>, age <dbl>, arm <chr>,  
## # sex <chr>, race <chr>, fu\_time <dbl>, fu\_stat <dbl>, ps <dbl>,  
## # hgb <dbl>, bmi <dbl>, alk\_phos <dbl>, ast <dbl>, mdquality\_s <dbl>,  
## # age\_ord <chr>, ethnicity <chr>, name <chr>, first\_name <chr>,  
## # last\_name <chr>, ae\_low\_wbc <dbl>, ae\_neuropathy <dbl>,  
## # ae\_diarrhea <dbl>, ae\_vomiting <dbl>, ae\_blood\_clot <dbl>, site <chr>,  
## # country <chr>, fu\_fct <fct>

Demographics

ggplot(mockdata, aes(age)) +  
 geom\_histogram(color = 'white',   
 fill = scico(1, palette = 'lapaz'),   
 bins = 20) +  
 labs(x = "Age",   
 y = "Count",   
 title = "Age Distribution in the \nCholangiocarcinoma Mock Study") +  
 scale\_y\_continuous(breaks = scales::pretty\_breaks())

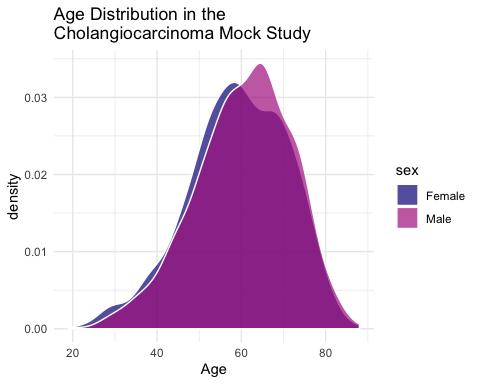


ggplot(mockdata, aes(age)) +  
 geom\_histogram(aes(fill = sex),  
 color = 'white',   
 bins = 20) +  
 labs(x = "Age",   
 y = "Count",   
 title = "Age Distribution in the \nCholangiocarcinoma Mock Study") +  
 facet\_wrap(vars(sex)) +  
 scale\_y\_continuous(breaks = scales::pretty\_breaks()) +  
 scale\_fill\_viridis\_d(option = "plasma", end = .4, alpha = .7, guide = FALSE)



# scale\_fill\_scico\_d(guide = FALSE, palette = 'lapaz', end = .4)

ggplot(mockdata, aes(age)) +  
 geom\_density(aes(fill = sex),  
 color = 'white') +  
 labs(x = "Age",   
 title = "Age Distribution in the \nCholangiocarcinoma Mock Study") +  
 scale\_fill\_viridis\_d(option = "plasma", end = .4, alpha = .7)



#scale\_fill\_scico\_d(palette = 'lapaz', end = .4, alpha = .7)

# a 1-way tabyl  
mockdata %>%   
 tabyl(sex) %>%   
 adorn\_totals("row") %>%  
 adorn\_pct\_formatting() %>%   
 knitr::kable()

|  |  |  |
| --- | --- | --- |
| sex | n | percent |
| Female | 583 | 38.9% |
| Male | 916 | 61.1% |
| Total | 1499 | 100.0% |

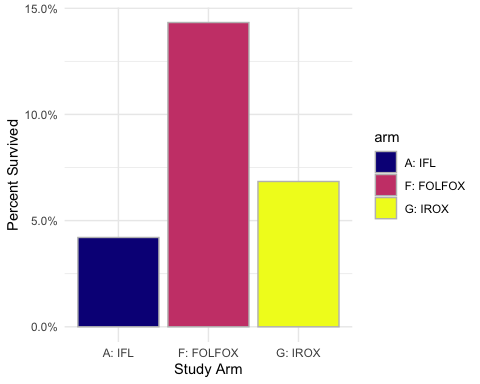
Follow-up stats

# followups, a 1-way tabyl  
mockdata %>%   
 tabyl(fu\_fct) %>%   
 adorn\_totals("row") %>%  
 adorn\_pct\_formatting() %>%   
 knitr::kable()

|  |  |  |
| --- | --- | --- |
| fu\_fct | n | percent |
| Lived | 143 | 9.5% |
| Died | 1356 | 90.5% |
| Total | 1499 | 100.0% |

# calculate proportion survived by arm  
prop\_surv <- mockdata %>%   
 count(arm, fu\_fct, name = "by\_surv", .drop = FALSE) %>%   
 add\_count(arm, wt = by\_surv, name = "arm\_total") %>%   
 mutate(prop = by\_surv/arm\_total) %>%   
 filter(fu\_fct == "Lived")

# barplot of percent survival by arm  
ggplot(prop\_surv, aes(x = arm, y = prop, fill = arm)) +  
 geom\_col(colour = "gray") +  
 labs(y= "Percent Survived", x= "Study Arm") +  
 scale\_fill\_viridis\_d(option = "plasma") +  
 scale\_y\_continuous(labels = scales::percent)



# a 2-way tabyl  
mockdata %>%   
 tabyl(fu\_fct, sex) %>%   
 adorn\_totals("row") %>% # can also do "col", or c("row", "col")  
 adorn\_percentages("row") %>%   
 adorn\_pct\_formatting(digits = 2) %>%   
 adorn\_ns() %>%   
 knitr::kable()

|  |  |  |
| --- | --- | --- |
| fu\_fct | Female | Male |
| Lived | 39.16% (56) | 60.84% (87) |
| Died | 38.86% (527) | 61.14% (829) |
| Total | 38.89% (583) | 61.11% (916) |

Arms

# a 2-way tabyl  
mockdata %>%   
 tabyl(arm, fu\_fct) %>%   
 adorn\_totals("row") %>% # can also do "col", or c("row", "col")  
 adorn\_percentages("row") %>%   
 adorn\_pct\_formatting(digits = 2) %>%   
 adorn\_ns() %>%   
 knitr::kable()

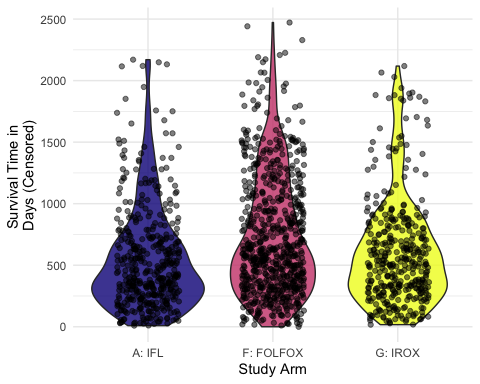
|  |  |  |
| --- | --- | --- |
| arm | Lived | Died |
| A: IFL | 4.21% (18) | 95.79% (410) |
| F: FOLFOX | 14.33% (99) | 85.67% (592) |
| G: IROX | 6.84% (26) | 93.16% (354) |
| Total | 9.54% (143) | 90.46% (1356) |

geom\_lollipop median survival time for each group

make bar/lollipop chart of proportions here (% survived)

Look at survival time now

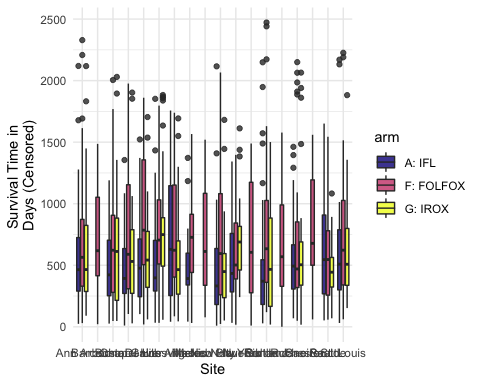
# distributions of survival time by arms (in days censored)  
ggplot(mockdata) +  
 aes(x=arm, y = fu\_time, fill = arm) +  
 geom\_violin(alpha = 0.8) +  
 geom\_jitter(width = 0.25, alpha = 0.5) +  
 labs(y= "Survival Time in \nDays (Censored)", x= "Study Arm") +  
 scale\_fill\_viridis\_d(option = "plasma", guide = FALSE)



#scale\_fill\_scico\_d(palette = 'lapaz', guide = FALSE)

sites

ggplot(mockdata,  
 aes(x = site, y = fu\_time, fill = arm)) +  
 geom\_boxplot(alpha = .8, position = position\_dodge2(preserve = "single")) +  
 labs(y= "Survival Time in \nDays (Censored)", x= "Site") +  
 scale\_fill\_viridis\_d(option = "plasma")



#scale\_fill\_scico\_d(palette = 'lapaz', guide = FALSE)

median\_days <- mockdata %>%   
 group\_by(arm, site) %>%   
 summarize(med\_fu = median(fu\_time))  
  
ggplot(median\_days,  
 aes(x = site, y = med\_fu, fill = arm)) +  
 geom\_linerange(aes(color = arm,   
 x = site,   
 ymin = 0,   
 ymax = med\_fu),  
 position = position\_dodge(width = .5)) +  
 geom\_point(colour = "black", shape = 21, position = position\_dodge(width = .5)) +  
 coord\_flip() +  
 scale\_fill\_viridis\_d(option = "plasma", end = .8) +  
 scale\_colour\_viridis\_d(option = "plasma", end = .8)

