Response and adverse events to chemotherapy: A mock study

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

## # A tibble: 1 x 1  
## site   
## <chr>   
## 1 Boston

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

Skim summary statistics  
n obs: 100  
n variables: 26

Variable type: character

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| variable | missing | complete | n | min | max | empty | n\_unique |
| age\_ord | 0 | 100 | 100 | 5 | 5 | 0 | 7 |
| arm | 0 | 100 | 100 | 6 | 9 | 0 | 3 |
| country | 0 | 100 | 100 | 3 | 3 | 0 | 1 |
| ethnicity | 0 | 100 | 100 | 8 | 25 | 0 | 4 |
| first\_name | 0 | 100 | 100 | 4 | 11 | 0 | 83 |
| last\_name | 0 | 100 | 100 | 3 | 16 | 0 | 97 |
| name | 0 | 100 | 100 | 9 | 24 | 0 | 100 |
| race | 0 | 100 | 100 | 8 | 14 | 0 | 4 |
| sex | 0 | 100 | 100 | 4 | 6 | 0 | 2 |
| site | 0 | 100 | 100 | 6 | 6 | 0 | 1 |

Variable type: factor

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| variable | missing | complete | n | n\_unique | top\_counts | ordered |
| fu\_fct | 0 | 100 | 100 | 2 | 2: 94, 1: 6, NA: 0 | FALSE |

Variable type: numeric

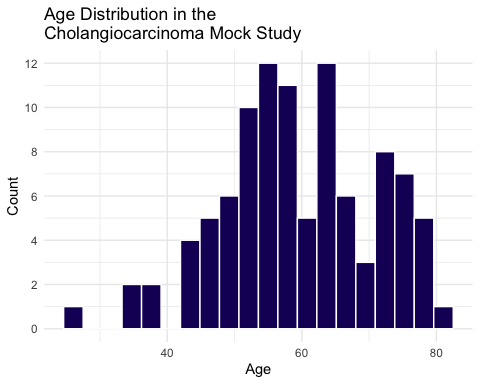
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | missing | complete | n | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| age | 0 | 100 | 100 | 59.41 | 11.29 | 27 | 51.75 | 58 | 67.25 | 82 | ▁▂▃▇▆▇▅▃ |
| alk\_phos | 19 | 81 | 100 | 169.21 | 116.17 | 41 | 93 | 124 | 221 | 635 | ▇▅▂▂▁▁▁▁ |
| ast | 19 | 81 | 100 | 39.59 | 31.92 | 9 | 20 | 26 | 47 | 189 | ▇▃▂▁▁▁▁▁ |
| blood\_clot | 0 | 100 | 100 | 0.04 | 0.2 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▁ |
| bmi | 6 | 94 | 100 | 26.55 | 4.92 | 16.84 | 23.21 | 26.03 | 29.66 | 39.29 | ▂▆▇▇▆▃▂▂ |
| case | 0 | 100 | 100 | 92589.45 | 238.94 | 92183 | 92401.5 | 92562 | 92793.25 | 93048 | ▅▅▇▇▅▆▃▅ |
| diarrhea | 0 | 100 | 100 | 0.18 | 0.39 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▂ |
| fu\_stat | 0 | 100 | 100 | 1.94 | 0.24 | 1 | 2 | 2 | 2 | 2 | ▁▁▁▁▁▁▁▇ |
| fu\_time | 0 | 100 | 100 | 611.66 | 434.49 | 26 | 252.5 | 573.5 | 830.5 | 2030 | ▇▅▇▅▂▁▁▁ |
| hgb | 19 | 81 | 100 | 12.6 | 1.8 | 9 | 11.3 | 12.4 | 13.9 | 17.1 | ▂▆▇▇▇▆▁▂ |
| low\_wbc | 0 | 100 | 100 | 0.13 | 0.34 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▁ |
| mdquality\_s | 11 | 89 | 100 | 0.88 | 0.33 | 0 | 1 | 1 | 1 | 1 | ▁▁▁▁▁▁▁▇ |
| neuropathy | 0 | 100 | 100 | 0.13 | 0.34 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▁ |
| ps | 19 | 81 | 100 | 0.47 | 0.55 | 0 | 0 | 0 | 1 | 2 | ▇▁▁▆▁▁▁▁ |
| vomiting | 0 | 100 | 100 | 0.18 | 0.39 | 0 | 0 | 0 | 0 | 1 | ▇▁▁▁▁▁▁▂ |

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

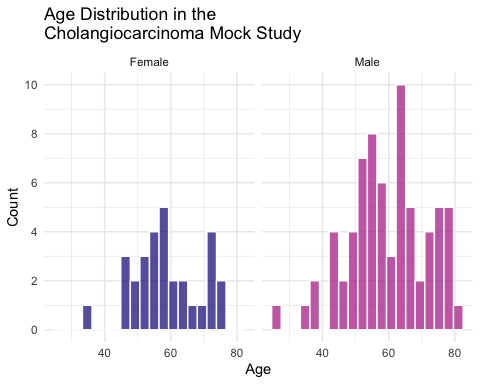
## # A tibble: 0 x 27  
## # … with 27 variables: case <dbl>, dupe\_count <int>, age <dbl>, arm <chr>,  
## # sex <chr>, race <chr>, fu\_time <dbl>, fu\_stat <dbl>, ps <dbl>,  
## # hgb <dbl>, bmi <dbl>, alk\_phos <dbl>, ast <dbl>, mdquality\_s <dbl>,  
## # age\_ord <chr>, ethnicity <chr>, name <chr>, first\_name <chr>,  
## # last\_name <chr>, low\_wbc <dbl>, neuropathy <dbl>, diarrhea <dbl>,  
## # vomiting <dbl>, 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())

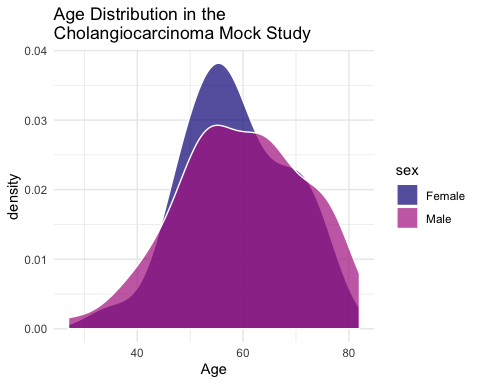


ggplot(mockdata, aes(age)) +  
 geom\_histogram(aes(fill = sex),  
 color = 'white',   
 bins = 20) +  
 labs(x = "Age",   
 y = "Count",   
 title = "Age Distribution in the \nCholangiocarcinoma Mock Study") +  
 facet\_wrap(vars(sex)) +  
 scale\_y\_continuous(breaks = scales::pretty\_breaks()) +  
 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 | 30 | 30.0% |
| Male | 70 | 70.0% |
| Total | 100 | 100.0% |

Follow-up stats

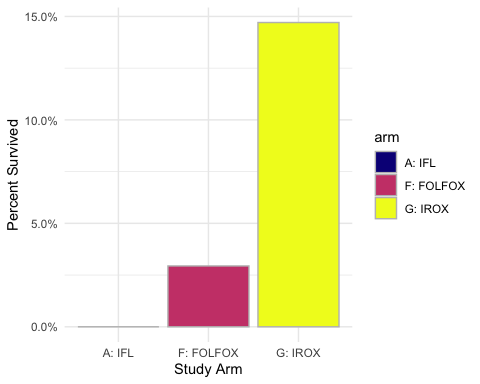
* 1 = lived
* 2 = died

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

|  |  |  |
| --- | --- | --- |
| fu\_fct | n | percent |
| 1 | 6 | 6.0% |
| 2 | 94 | 94.0% |
| Total | 100 | 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 == 1)

# 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 |
| 1 | 16.67% (1) | 83.33% (5) |
| 2 | 30.85% (29) | 69.15% (65) |
| Total | 30.00% (30) | 70.00% (70) |

Arms (1 = lived, 2 = died)

# 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()

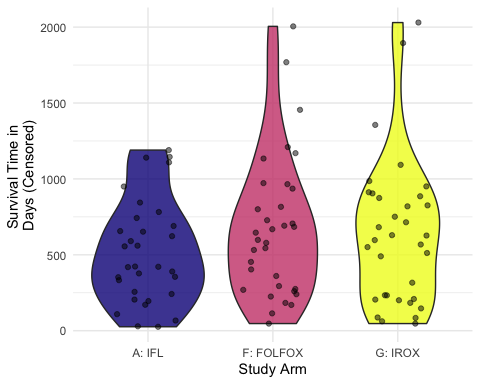
|  |  |  |
| --- | --- | --- |
| arm | 1 | 2 |
| A: IFL | 0.00% (0) | 100.00% (32) |
| F: FOLFOX | 2.94% (1) | 97.06% (33) |
| G: IROX | 14.71% (5) | 85.29% (29) |
| Total | 6.00% (6) | 94.00% (94) |

geom\_lollipop median survival time for each group

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

Look at survival time now

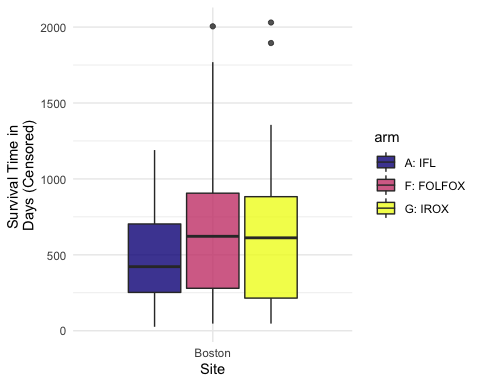
# distributions of survival time by arms (in days censored)  
ggplot(mockdata) +  
 aes(x=arm, y = fu\_time, fill = arm) +  
 geom\_violin(alpha = 0.8) +  
 geom\_jitter(width = 0.25, alpha = 0.5) +  
 labs(y= "Survival Time in \nDays (Censored)", x= "Study Arm") +  
 scale\_fill\_viridis\_d(option =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



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