

```

library(survival)

## Loading required package: splines
load("cpvs.20150119.RData")

sapply(data.clin, function(d) c("days_to_death", "days_to_last_followup", "vital_status") %in% colnames(d))

##      acc blca brca cesc coad dlbc esca  gbm hnscc kich kirc kirp lam1 lgg
## [1,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [3,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##      lihc luad lusc meso  ov paad prad read sarc skcm stad thca ucec ucs
## [1,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [2,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [3,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE

data.clin.merged = do.call(rbind, lapply(data.clin, function(d) apply(d[,c("bcr_patient_barcode", "days_to_death", "days_to_last_followup", "vital_status"), 2, function(x) x)])))
data.clin.merged = data.frame(data.clin.merged, stringsAsFactors = FALSE)
data.clin.merged$days_to_death = as.numeric(data.clin.merged$days_to_death)

## Warning: NAs introduced by coercion

data.clin.merged$days_to_initial_pathologic_diagnosis = as.numeric(data.clin.merged$days_to_initial_pathologic_diagnosis)

## Warning: NAs introduced by coercion

data.clin.merged$days_to_last_followup = as.numeric(data.clin.merged$days_to_last_followup)

## Warning: NAs introduced by coercion

data.clin.merged$vital_status = as.factor(data.clin.merged$vital_status)
data.clin.merged$cancer = factor(rep(names(data.clin), sapply(data.clin, nrow)))
data.clin.merged$event = data.clin.merged$vital_status %in% c("Dead", "DECEASED")

data.clin.merged$time = NA
data.clin.merged$time[data.clin.merged$event] = data.clin.merged$days_to_death[data.clin.merged$event]
data.clin.merged$time[!data.clin.merged$event] = data.clin.merged$days_to_last_followup[!data.clin.merged$event]

data.clin.merged = data.clin.merged[!is.na(data.clin.merged$time) & !is.na(data.clin.merged$event),]

library(plyr)
fits = dplyr::group_by(data.clin.merged, cancer) %>% summarise(fit = survreg(Surv(days_to_death, days_to_last_followup) ~ 1, dist = "logistic"))

sort(sapply(fits, function(f) f$scale) / sapply(fits, coef))

##      prad  thca  read  brca  ucec  kich  stad  kirp  dlbc  lgg
## 0.1226 0.2120 0.2413 0.2717 0.2798 0.2890 0.2916 0.3026 0.3037 0.3223
##      paad  coad  luad  kirc  ov  cesc  lihc  lusc  blca  sarc
## 0.3249 0.3383 0.3519 0.3561 0.3612 0.3647 0.4011 0.4091 0.4259 0.4306
##      esca  hnscc  skcm  acc  ucs  gbm  meso  lam1
## 0.4438 0.4445 0.4449 0.4507 0.5047 0.5162 0.6023 0.6408

rel_iqrs = dplyr::group_by(data.clin.merged, cancer) %>% summarise(fit = survfit(Surv(days_to_death, days_to_last_followup) ~ 1),
  qs = approx(fit$surv, fit$time, c(0.25, 0.5, 0.75))$y,
  (qs[1]-qs[3]) / qs[2])

sort(unlist(rel_iqrs))

```

```
## dlbc paad ov stad lgg gbm luad lihc skcm lusc
## 0.5399 0.8401 0.9979 1.0353 1.0976 1.1012 1.3945 1.5896 1.6648 1.6710
## hnscl ucs meso lamll
## 1.7267 2.1334 2.8130 2.9566

# Right. So PDAC is basically one of the most consistent cancers re: survival time.
# Well there goes that argument.
```

```
# Top 10 from AIHW2014:
# Prostate (C61) 19,993
# Colorectal (C18C20) 15,151
# Breast in females (C50) 14,465
# Melanoma of the skin (C43) 11,570
# Lung (C33C34) 10,511
# Non-Hodgkin lymphoma (C82C85) 4,631
# Kidney (C64) 2,847
# Pancreas (C25) 2,748
# Bladder (C67) 2,404
# Uterus (C54C55) 2,238

# Stomach (C16) 2,093

library(ggplot2)

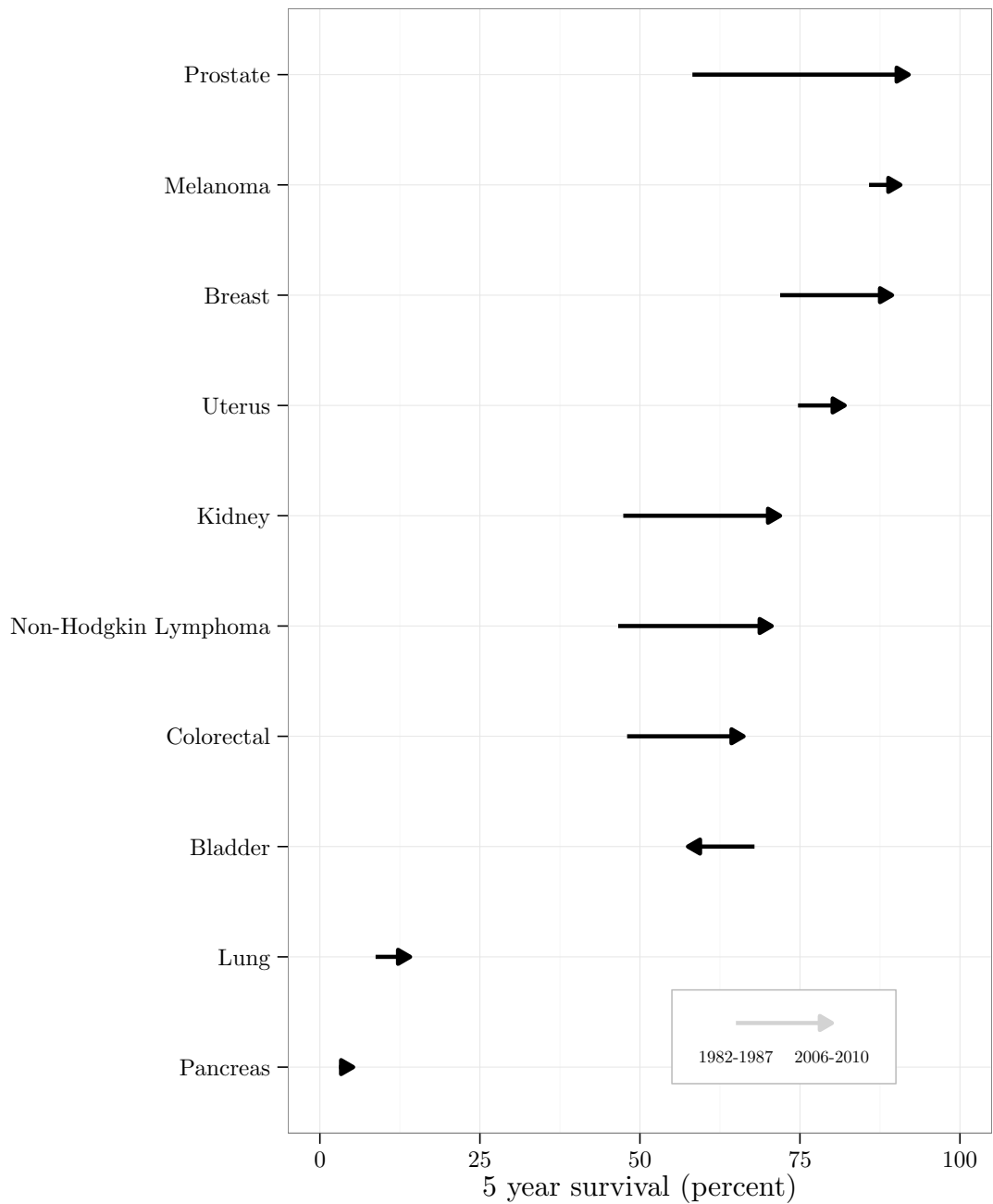
## Loading required package: methods

library(grid)

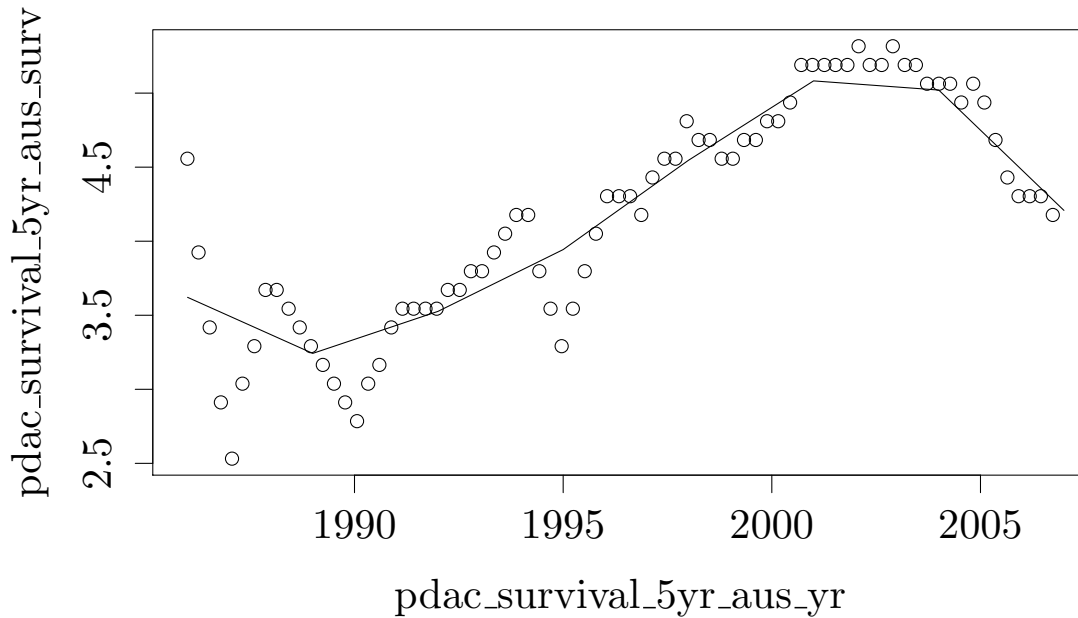
rm(list = ls())
cancer = c("Prostate", "Colorectal", "Breast", "Melanoma", "Lung", "Non-Hodgkin Lymphoma", "Pancreas", "Kidney", "Bladder", "Uterus", "Stomach")
incidence_2011 = c(19993, 15151, 14465, 11570, 10511, 4631, 2847, 2748, 2404, 2238, 2093)
survival_5yr_1982 = c(58.2, 48.0, 71.9, 85.8, 8.7, 46.6, 47.1, 46.6, 46.6, 46.6, 46.6)
survival_5yr_2010 = c(92.0, 66.2, 89.4, 90.7, 14.1, 70.6, 71.9, 70.6, 70.6, 70.6, 70.6)

# survival_historical = data.frame(cancer = rep(cancer, 2), incidence = rep(incidence_2011, 2), incidence_rank = rep(rank(incidence_2011), 2))
survival_historical = data.frame(cancer = cancer, incidence = incidence_2011, incidence_rank = rank(incidence_2011))

survival_historical$cancer = ordered(as.character(survival_historical$cancer), levels = as.character(survival_historical$cancer))
ggplot(survival_historical, aes(x = cancer, xend = cancer, y = surv1, yend = surv2)) +
  coord_flip() +
  geom_segment(arrow = arrow(length = unit(0.2, "cm"), type = "closed"), lwd = 1.5) +
  ylab("5 year survival (percent)") + xlab("") + theme_bw() + ylim(0, 100) +
  annotate("rect", xmin = 0.85, xmax = 1.7, ymin = 55, ymax = 90, fill = "white", colour = "grey") +
  annotate("segment", x = 1.4, xend = 1.4, y = 65, yend = 80, colour = "lightgrey", lwd = 1.5, arrow = FALSE) +
  annotate("text", x = 1.1, y = 65, label = "1982-1987", cex = 2.5) +
  annotate("text", x = 1.1, y = 80, label = "2006-2010", cex = 2.5)
```



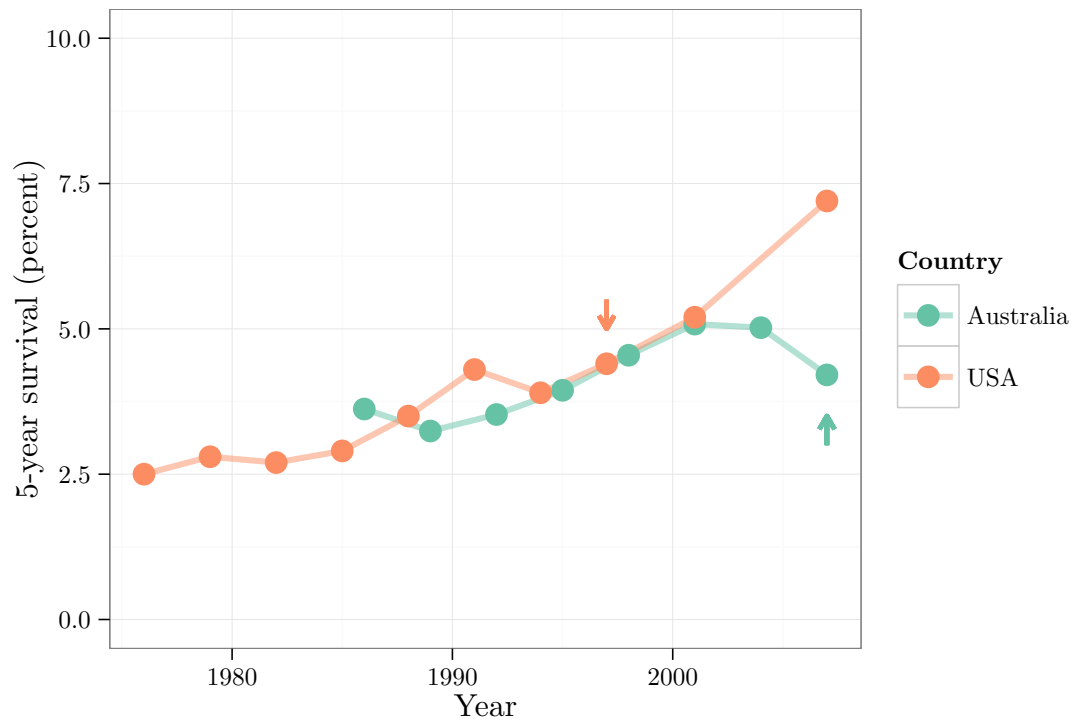
```
survival_historical$cancer = ordered(as.character(survival_historical$cancer), levels = as.character(survival_historical$cancer))
ggplot(survival_historical, aes(x = cancer, xend = cancer, y = surv1, yend = surv2)) +
  coord_flip() +
  geom_segment(arrow = arrow(length = unit(0.2, "cm"), type = "closed"), lwd = 1.5) +
  ylab("5 year survival (percent)") + xlab("") + theme_bw() + ylim(0, 100)
```

```
pdac_survival_5yr_aus = data.frame(Year = pdac_survival_5yr_aus_yr2, Survival = pdac_survival_5yr_aus_surv)

# Digitized data from NCI SEER:
pdac_survival_5yr_usa_year = c(seq(1976, 1997, 3), 2001, 2007)
pdac_survival_5yr_usa_surv = c(2.5, 2.8, 2.7, 2.9, 3.5, 4.3, 3.9, 4.4, 5.2, 7.2)
pdac_survival_5yr_usa = data.frame(Year = pdac_survival_5yr_usa_year, Survival = pdac_survival_5yr_usa_surv)

pdac_survival_5yr = rbind(cbind(pdac_survival_5yr_aus, Country = "Australia"), cbind(pdac_survival_5yr_usa, Country = "USA"))
ggplot(pdac_survival_5yr, aes(x = Year, y = Survival, colour = Country)) +
  geom_line(lwd = 2, alpha = 0.5) +
  geom_point(cex = 4) +
  ylim(0, 10) +
  theme_bw() +
  geom_segment(x = 1997, xend = 1997, y = 5.5, yend = 5, colour = "#fc8d62", alpha = 0.5, lwd = 1) +
  geom_segment(x = 2007, xend = 2007, y = 3, yend = 3.5, colour = "#66c2a5", alpha = 0.5, lwd = 1) +
  ylab("5-year survival (percent)") +
  scale_colour_brewer(palette = "Set2")
```



Surgery (adjuvant): Gemcitabine / 5-FU / capecitabine (w/ or w/o radio)
 # Metastatic: FOLFIRINOX / Gem-nabpac / Gem-erlotinib / gem / 5-FU
 # FOLFIRINOX 2010
 # Gem 1996 (UK) - 1997 (USA) - 2007 (Aus)