1.Loading the required packages

In [1]:

```
library(RCurl)
library(tidyr)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(ggthemes)
```

```
Warning message:
"package 'RCurl' was built under R version 3.6.3"
Attaching package: 'tidyr'

The following object is masked from 'package:RCurl':

    complete

Registered S3 methods overwritten by 'ggplot2':
  method         from
  [.quosures     rlang
  c.quosures     rlang
  print.quosures rlang

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

Registered S3 method overwritten by 'rvest':
  method            from
  read_xml.response xml2
-- Attaching packages --------------------------------------- tidyverse 1.
2.1 --
v tibble  2.1.1     v purrr   0.3.2
v readr   1.3.1     v stringr 1.4.0
v tibble  2.1.1     v forcats 0.4.0
-- Conflicts ------------------------------------------ tidyverse_conflict
s() --
x tidyr::complete() masks RCurl::complete()
x dplyr::filter()   masks stats::filter()
x dplyr::lag()      masks stats::lag()
Warning message:
"package 'ggthemes' was built under R version 3.6.3"
```

2. Importing the dataset and basic data viz

In [2]:

```
netflix <- read.csv (text = getURL("https://raw.githubusercontent.com/satwikshankar/Projects/master/input.csv"))

# Formating the date_added column to workable format

netflix$date_added <- as.Date(netflix$date_added, format = "%B %d, %Y")

# Basic data distribution

head(netflix)
glimpse(netflix)
```

show_id	type	title	director	cast	country	date_added	release_year	ra
81145628	Movie	Norm of the North: King Sized Adventure	Richard Finn, Tim Maltby	Alan Marriott, Andrew Toth, Brian Dobson, Cole Howard, Jennifer Cameron, Jonathan Holmes, Lee Tockar, Lisa Durupt, Maya Kay, Michael Dobson	United States, India, South Korea, China	2019-09-09	2019	F
80117401	Movie	Jandino: Whatever it Takes		Jandino Asporaat	United Kingdom	2016-09-09	2016	
70234439	TV Show	Transformers Prime		Peter Cullen, Sumalee Montano, Frank Welker, Jeffrey Combs, Kevin Michael Richardson, Tania Gunadi, Josh Keaton, Steve Blum, Andy Pessoa, Ernie Hudson, Daran Norris, Will Friedle	United States	2018-09-08	2013	Y7
80058654	TV Show	Transformers: Robots in Disguise		Will Friedle, Darren Criss, Constance Zimmer, Khary Payton, Mitchell Whitfield, Stuart Allan, Ted McGinley, Peter Cullen	United States	2018-09-08	2016	TV

show_id	type	title	director	cast	country	date_added	release_year	ra
80125979	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins, Keith Powers, Alicia Sanz, Jake Borelli, Kid Ink, Yousef Erakat, Rebekah Graf, Anne Winters, Peter Gilroy, Patrick Davis	United States	2017-09-08	2017	TV
80163890	TV Show	Apaches		Alberto Ammann, Eloy Azorín, Verónica Echegui, Lucía Jiménez, Claudia Traisac	Spain	2017-09-08	2016	

```
Observations: 6,234
Variables: 12
$ show_id      <int> 81145628, 80117401, 70234439, 80058654, 80125979, 801
6...
$ type         <fct> Movie, Movie, TV Show, TV Show, Movie, TV Show, Movi
e,...
$ title        <fct> Norm of the North: King Sized Adventure, Jandino: Wha
t...
$ director     <fct> "Richard Finn, Tim Maltby", "", "", "", "Fernando Leb
r...
$ cast         <fct> "Alan Marriott, Andrew Toth, Brian Dobson, Cole Howar
d...
$ country      <fct> "United States, India, South Korea, China", "United K
i...
$ date_added   <date> 2019-09-09, 2016-09-09, 2018-09-08, 2018-09-08, 2017
-...
$ release_year <int> 2019, 2016, 2013, 2016, 2017, 2016, 2014, 2017, 2017,
...
$ rating       <fct> TV-PG, TV-MA, TV-Y7-FV, TV-Y7, TV-14, TV-MA, R, TV-M
A,...
$ duration     <fct> 90 min, 94 min, 1 Season, 1 Season, 99 min, 1 Season,
...
$ listed_in    <fct> "Children & Family Movies, Comedies", "Stand-Up Comed
y...
$ description  <fct> "Before planning an awesome wedding for his grandfath
e...
```

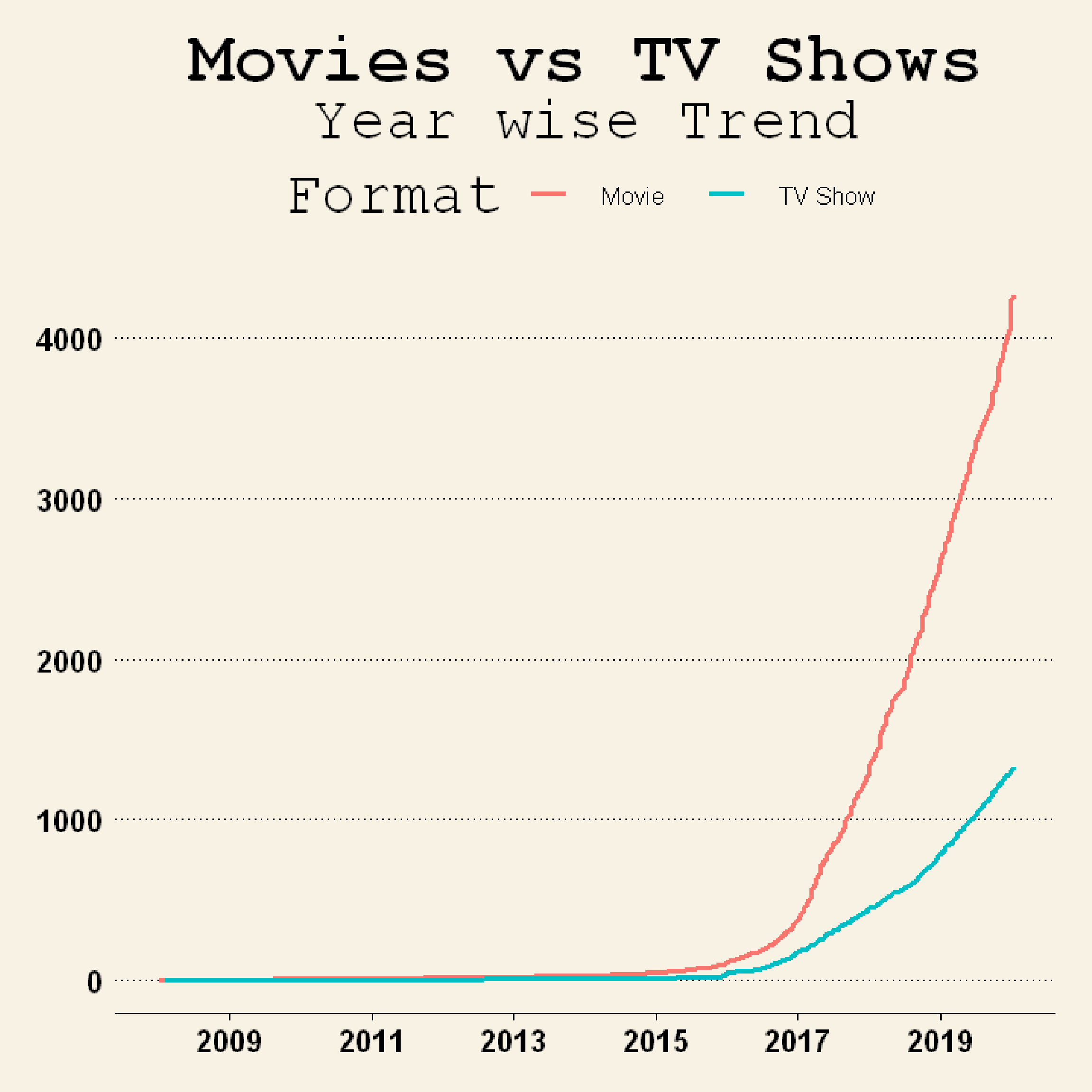
3. Year wise trend

In [3]:

```
options(repr.plot.width = 6, repr.plot.height = 6)

netflix_year <- netflix %>%
    group_by(date_added,type) %>%
    summarise(shows_added = n()) %>%
    ungroup() %>%
    group_by(type) %>%
    mutate(Total_Number_of_Shows = cumsum(shows_added))


head(netflix_year)


netflix_year %>%
ggplot(aes(x = date_added, y = Total_Number_of_Shows, color = type)) +
geom_line(size = 1) +
theme_wsj() +
theme(plot.title= element_text(hjust = 0.5), plot.subtitle = element_text(hjust = 0.5))
+
scale_x_date(date_breaks = '2 years', date_labels = "%Y") +
labs(color = "Format", title="Movies vs TV Shows", subtitle = "Year wise Trend",  y =
"No. of Shows", x = "Year")
```

date_added	type	shows_added	Total_Number_of_Shows
2008-01-01	Movie	1	1
2008-02-04	TV Show	1	1
2009-05-05	Movie	1	2
2009-11-18	Movie	1	3
2010-11-01	Movie	1	4
2011-05-17	Movie	1	5

```
Warning message:
"Removed 2 rows containing missing values (geom_path)."
```

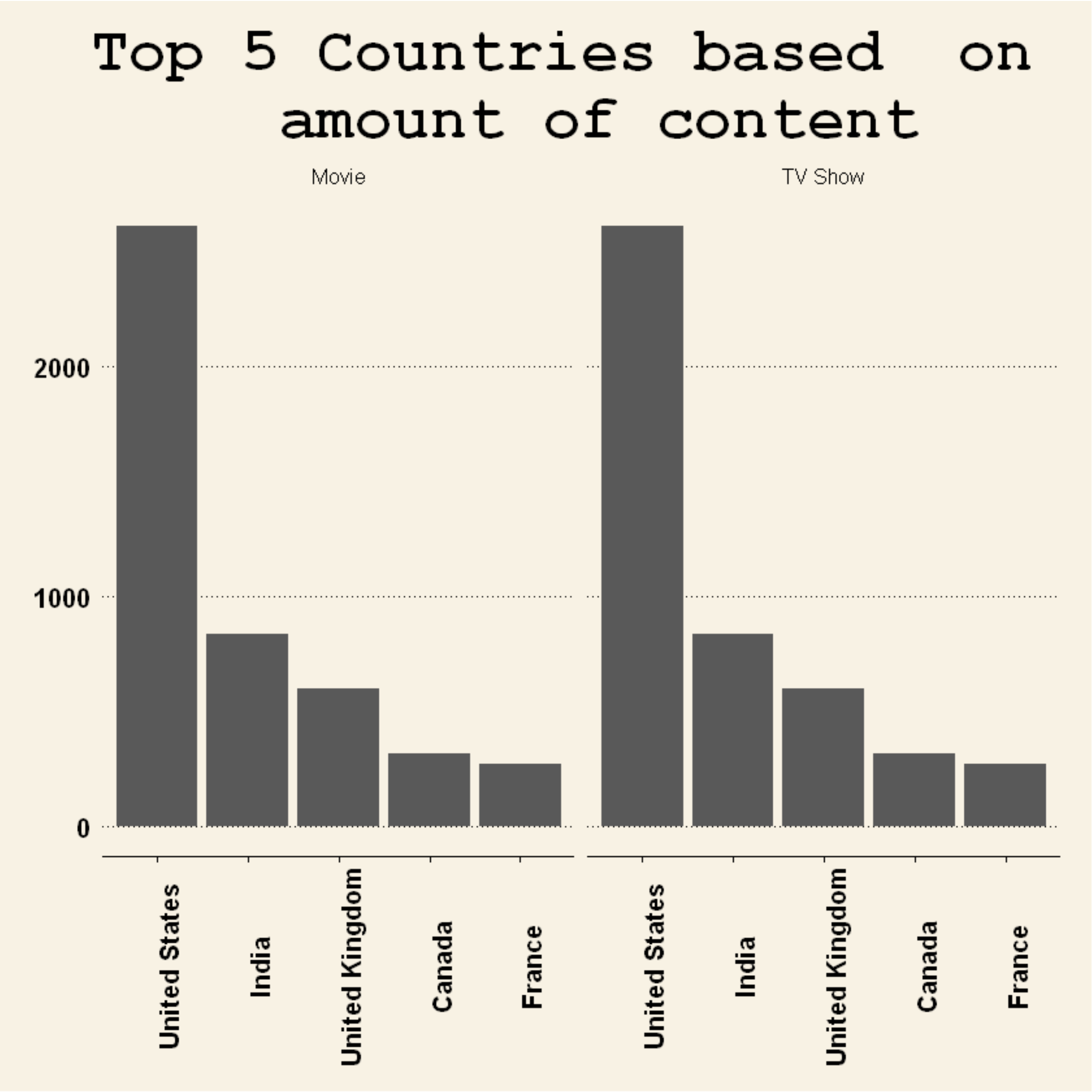


4. Countrywise content availability (Top 5)

In [4]:

```
options(repr.plot.width = 7, repr.plot.height = 7)


netflix_country <- netflix %>%
    mutate(country = strsplit(as.character(country), ",")) %>%
    unnest(country) %>%
    mutate(country = trimws(country, which = c("left"))) %>%
    group_by(country)%>%
    add_tally()

netflix_country <- netflix_country %>%
    select(country,n,type) %>%
    unique() %>%
    arrange(desc(n))

netflix_country_top5 <- netflix_country[1:10,]

head(netflix_country_top5)

ggplot( netflix_country_top5, aes(x = fct_reorder(country, n, .desc = TRUE), y = n))+
    geom_bar(stat = "identity")+
    facet_wrap(~type)+
    theme_wsj()+
    theme(plot.title= element_text(hjust = 0.5))+
    theme(plot.title= element_text(hjust = 0.5), axis.text.x = element_text(angle = 90
), legend.position = 'none') +
    labs(title="Top 5 Countries based  on \n amount of content", y = "Content", x = "Co
untry")
```

country	n	type
United States	2610	Movie
United States	2610	TV Show
India	838	Movie
India	838	TV Show
United Kingdom	602	Movie
United Kingdom	602	TV Show

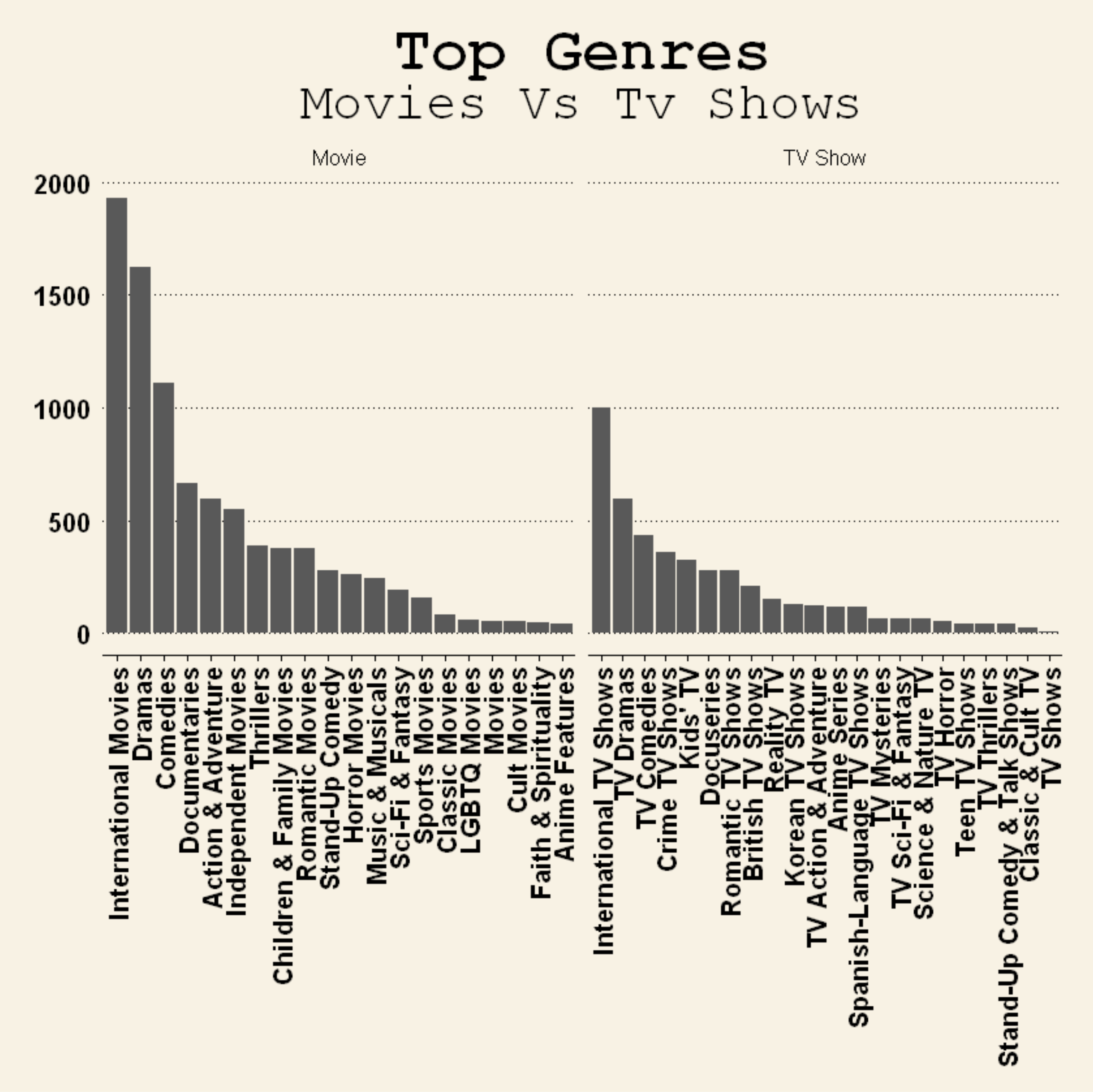

Top 5 Countries based on amount of content
Movie
TV Show
2000
1000
0
United States
India
United Kingdom
Canada
France
United States
India
United Kingdom
Canada
France

5. Genre distribution

In [5]:

```
options(repr.plot.width = 7, repr.plot.height = 7)

show_categories <- netflix %>%
    mutate(genre = strsplit(as.character(listed_in), ",")) %>%
    unnest(genre) %>%
    mutate(genre = trimws(genre, which = "left")) %>%
    group_by(type, genre) %>%
    summarise(count = n()) %>%
    unique()


head(show_categories)

show_categories %>%
    ggplot(aes(x = fct_reorder(genre, count, .desc = TRUE), y = count)) +
    geom_bar(stat = "identity") +
    scale_x_discrete() +
    facet_wrap(~type, scales = 'free_x') +
    theme_wsj() +
    theme(plot.title= element_text(hjust = 0.5), plot.subtitle = element_text(hjust =
0.5))+
    labs(title="Top Genres", subtitle = "Movies Vs Tv Shows", x="Genres", y = "Count")
+
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

type	genre	count
Movie	Action & Adventure	597
Movie	Anime Features	45
Movie	Children & Family Movies	378
Movie	Classic Movies	84
Movie	Comedies	1113
Movie	Cult Movies	55



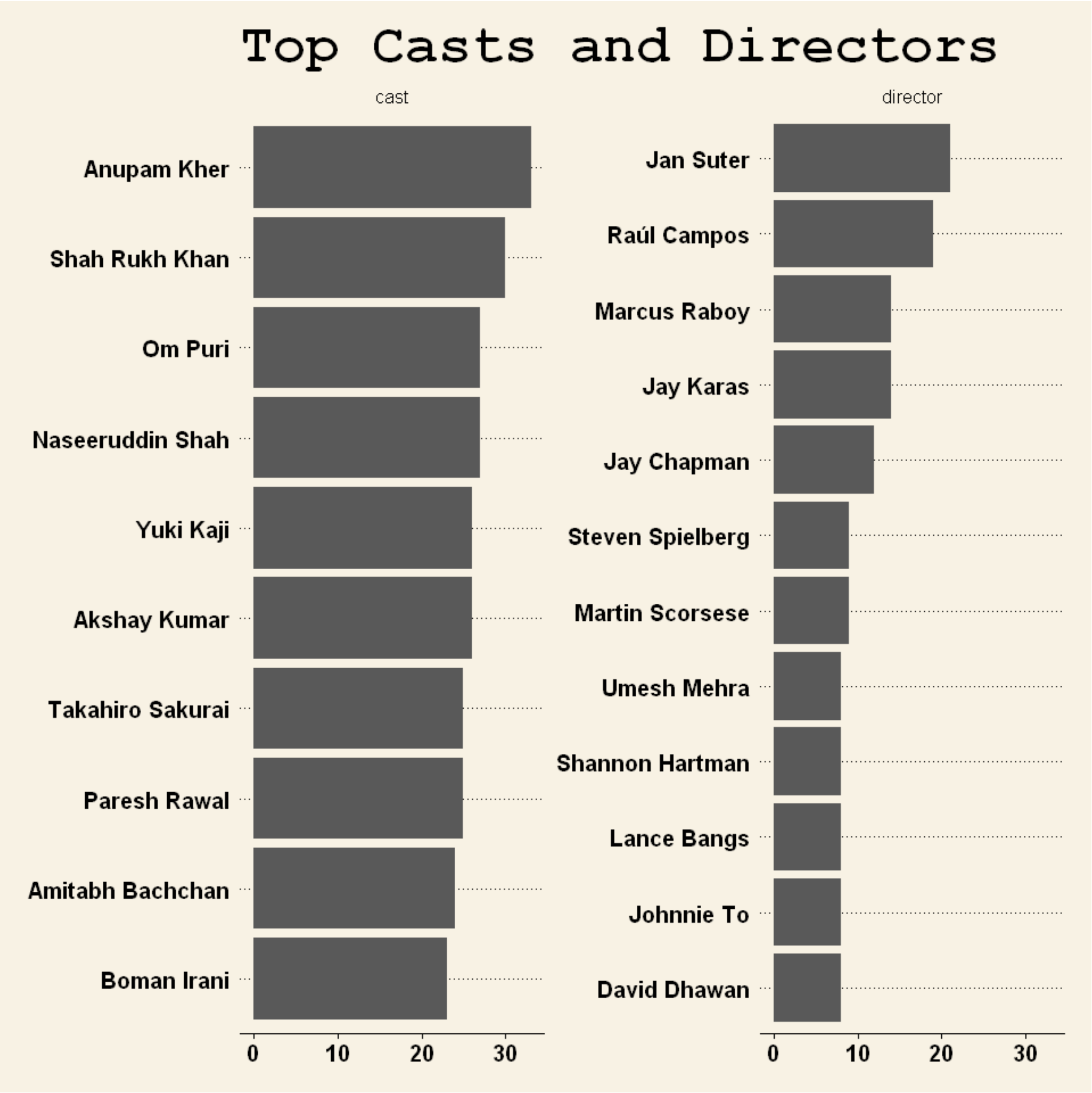
6. Top Actors and Directors based on Content Quantity

In [6]:

```
options(repr.plot.width = 8, repr.plot.height = 8)


netflix_cast <- netflix %>%
    select(c("cast","director"))%>%
    gather(key = role, value = name, cast, director) %>%
    filter(name != "") %>%
    mutate(name = strsplit(as.character(name), ",")) %>%
    unnest(name) %>%
    mutate(name = trimws(name, which = "left")) %>%
    group_by(name,role) %>%
    summarise(count = n())

head(netflix_cast)

netflix_cast %>%
    group_by(role) %>%
    top_n(10,count) %>%
    ungroup() %>%
    ggplot(aes(x = reorder(name,count), y = count)) +
    geom_bar(stat = 'identity') +
    coord_flip()+
    facet_wrap(role~., scales = 'free_y') +
    theme_wsj()+
    labs(title="Top Casts and Directors")
```

```
Warning message:
"attributes are not identical across measure variables;
they will be dropped"
```

name	role	count
2 Chainz	cast	1
4Minute	cast	1
50 Cent	cast	3
A-ra Go	cast	1
A Boogie Wit tha Hoodie	cast	1
A. L. Vijay	director	2


Top Casts and Directors
cast
Anupam Kher
Shah Rukh Khan
Om Puri
Naseeruddin Shah
Yuki Kaji
Akshay Kumar
Takahiro Sakurai
Paresh Rawal
Amitabh Bachchan
Boman Irani
0
10
20
30
director
Jan Suter
Raúl Campos
Marcus Raboy
Jay Karas
Jay Chapman
Steven Spielberg
Martin Scorsese
Umesh Mehra
Shannon Hartman
Lance Bangs
Johnnie To
David Dhawan
0
10
20
30