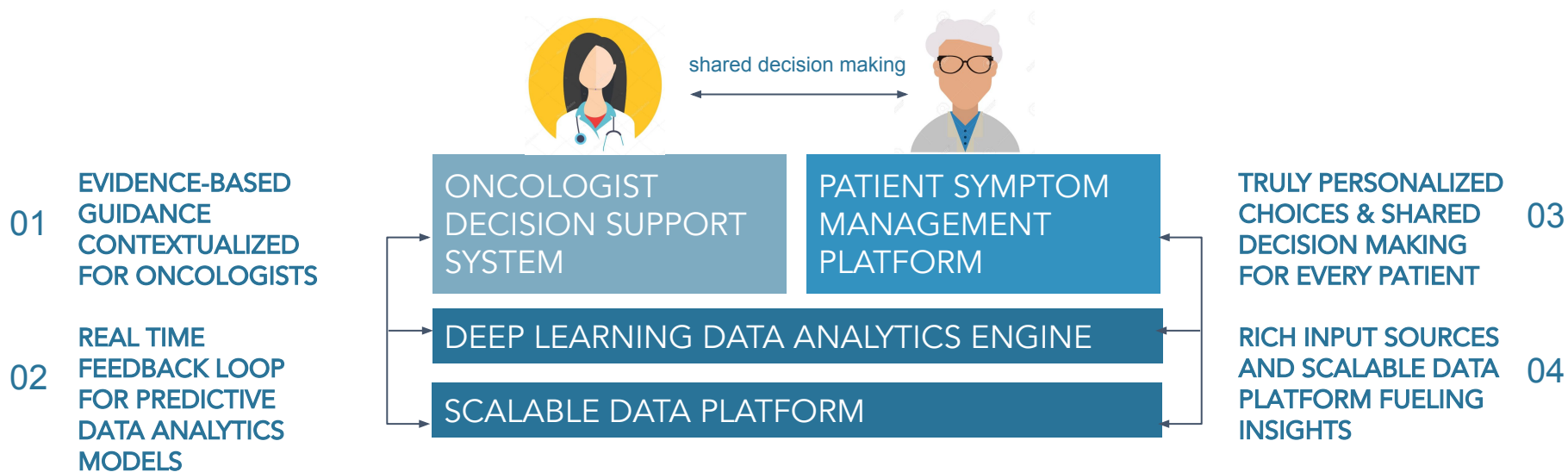# RŌNIN

Cancer intelligence for the N of 1

R-Ladies Meetup | Clara Oromendia | August 2019

WE'RE BUILDING A CANCER DECISION SUPPORT SYSTEM THAT DELIVERS INSIGHTS AND EFFICIENCIES AT THE POINT OF CARE

shared decision making

01 EVIDENCE-BASED GUIDANCE CONTEXTUALIZED FOR ONCOLOGISTS

02 REAL TIME FEEDBACK LOOP FOR PREDICTIVE DATA ANALYTICS MODELS

ONCOLOGIST DECISION SUPPORT SYSTEM

PATIENT SYMPTOM MANAGEMENT PLATFORM

DEEP LEARNING DATA ANALYTICS ENGINE

SCALABLE DATA PLATFORM

03 TRULY PERSONALIZED CHOICES & SHARED DECISION MAKING FOR EVERY PATIENT

04 RICH INPUT SOURCES AND SCALABLE DATA PLATFORM FUELING INSIGHTS

Source of Data: Clinical Trials

Many
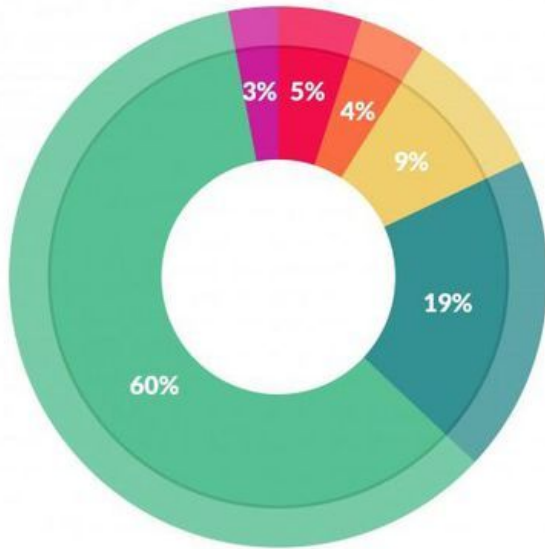**variables**,
few trials?

Few variables,
many **trials**?

The 80/20 data science dilemma

Most data scientists spend only 20 percent of their time on actual data analysis and 80 percent of their time finding, cleaning, and reorganizing huge amounts of data, which is an inefficient data strategy

What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html

Adverse Events Data

Adverse Events Data

For each adverse event, we want only 5 things:
1. Patient
2. Day started
3. Day ended
4. Type of event
5. Severity grade

	patient	start_day	end_day	name	grade
1	A	10	12	headache	1
2	B	32	33	fatigue	3
3	C	7	27	nausea	2

Trial 1: Amgen

✓ Dataset contains all information

✓ Labels provide description

Need to calculate end day

```r
# Import
dat_amgen_raw <- haven::read_sas("Prostat_Amgen_2006_155_aae.sas7bdat")
# Munge
dat_amgen_raw %>%
  select(SUBJID, AEDECOD, AESTDY, AEDURD, AESER, AETOXGR,AECONT) %>%
  mutate(end_day = AESTDY + AEDURD) %>%
  rename(patient = SUBJID,
         start_day = AESTDY,
         name = AEDECOD,
         grade = AETOXGR) ->
  dat_amgen

# Combine
dat_ideal %>%
  full_join(dat_amgen) ->
  dat_full
```

patient   name   start_day   end_day   grade

	SUBJID Subject Identifier for the Study	AEDECOD Dictionary–Derived Term	AESTDY Study Day of Start of Adverse Event	AEDURD Duration of Event in Days	AESER Serious Event	AETOXGR Standard Toxicity Grade	AECON Advers
1	001163	Acquired phimosis	471	5	N	02	
2	001163	Actinic keratosis	581	1	N	02	
3	001163	Actinic keratosis	581	1	N	02	
4	001163	Cataract	1050	1	N	01	
5	001163	Cataract	915	11	N	01	
6	001163	Contusion	133	10	N	01	

Trial 2: AstraZeneca

✓ Dataset contains all information

✓ Labels provide description

Need to filter to Adverse Events only

```{r}
# Import
dat_astraZe_raw <- haven::read_sas("Prostat_AstraZe_2008_103_r_aevbb.sas7bdat")
# Munge
dat_astraZe_raw %>%
  rename(patient = SUBJ,
         start_day = AESDYRND,
         end_day = AEEDYRND,
         name = PT_NAME,
         grade = CTCGMAX) %>%
  filter(AEEVNT01 == 0) %>%
  select(patient, start_day, end_day, name, grade) ->
  dat_astraZe

# Combine
dat_ideal %>%
  full_join(dat_amgen) %>%
  full_join(dat_astraZe) ->
  dat_full
```

patient	start_day	end_day	name	grade	
SUBJ Random patient identifier	AESDYRND AE start study day – randomisation	AEEDYRND AE end study day – randomisation	PT_NAME MedDRA Preferred Term Name	CTCGMAX Maximum CTC grade	AEEVNT01 Adverse Event
39	359	68	88	Urinary tract infection	1	1
40	359	68	71	Vomiting	1	1
41	359	102	NA	Insomnia	1	1
42	359	145	206	Tendonitis	1	1
43	286	100	101	Inguinal hernia	3	1
44	45	NA	NA			0

Only 250 more trials to go!

(And 100 more variables
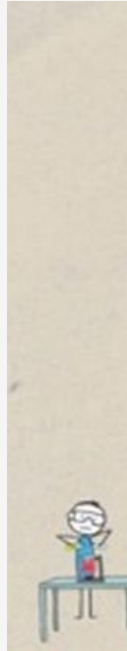
Banging my head vigorously against my desk is surely imminent.

your ecards
someecards.com

Time consuming & repetitive process -> find a better option

Better option: tagging

Amgen Meta

AstraZeneca Meta

meta_Prostat_Amgen_2006_155_aae.sas7bdat.csv

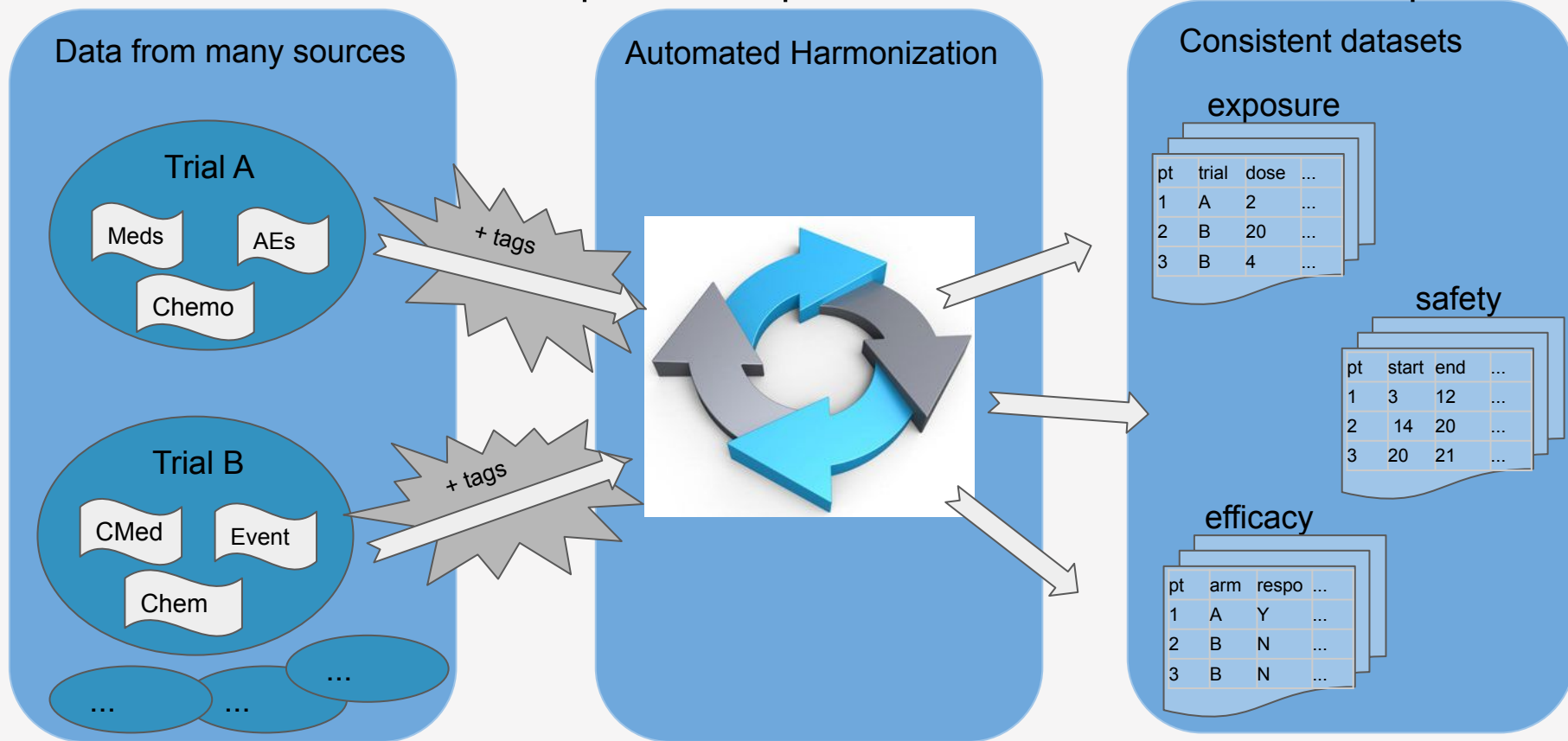| name	label	class	tag	head	unique_vals
STUDYID	Study Identifier	character		20050147 20050147 200501...	20050147
SUBJID	Subject Identifier for the Study	character	patient	001163 001163 001163 0011...	001163 00073...
GEOREG	Geographic Region	character		NORTH AMERICA NORTH A...	NORTH AMERI...
GEOREGCD	Geographic Region Code	character		NA NA NA NA NA NA	NA OT EU LA
TRTA	Actual Treatment Received	character		Controlled Arm Controlled Ar...	Controlled Arm
AGEY	Age at Enrollment in Years	numeric		86 86 86 86 86 86	86 77 72 87 8...
SEX	Sex	character		M M M M M M	M
RACE	Race	character		WHITE OR CAUCASIAN WHIT...	WHITE OR CA...
RACECD	Race Code	character		01 01 01 01 01 01	01 02 03 88 04
AEDECOD	Dictionary-Derived Term	character	name	Acquired phimosis Actinic ke...	Acquired phim...
AEBODSYS	Body System or Organ Class	character		Reproductive system and bre...	Reproductive s...
AELLT	Adverse Event Low Level Term	character		Paraphimosis Actinic keratosi...	Paraphimosis...
AEHLT	Adverse Event High Level Term	character		Penile disorders NEC (excl er...	Penile disorder...
AEHLGT	Adverse Event High Level Group Term	character		Penile and scrotal disorders (...	Penile and scr...
AESTDY	Study Day of Start of Adverse Event	numeric	start_day	471 581 581 187 NA 1050	471 581 187...
AEDURD	Duration of Event in Days	numeric	duration	5 1 1 NA NA 1	5 1 NA 11 1...
AECONT	Adverse Event Continuing	character			Y
AETOXGR	Standard Toxicity Grade	character	grade	02 02 02 02 02 01	02 01 04 03 05
AESER	Serious Event	character		N N N N Y N	N Y
AEREL	Causality	character		NOT RELATED NOT RELATED...	NOT RELATED...
AEACTLST	List of Action Code	character		88 88 88 01 04,88 88	88 01 04,88 0...
AESDTH	Results in Death	character		N N N N N N	N Y
AESLIFE	Is Life Threatening	character		N N N N N N	N Y

meta_Prostat_AstraZe_2008_103_r_aevbb.sas7bdat.csv

name	label	class	tag	head	unique_vals
AESER	Serious AE	character		0 0 1 0 1 0	0 1
SUBJ	Random patien...	numeric	patient	252 265 265 407 407 455	252 265 407 455 522 61 29...
AEACTIP	Action Taken, l...	character		0 0 0 0 0 0	0 3 2
CTCGMAX	Maximum CTC...	character	grade	2 1 2 2 3 1	2 1 3 5 4
AETREAT	Subject Recd T...	character		0 1 1 0 0 0	0 1
AECAUDIS	AE Caused Su...	character			0 1
AECAUS	Reasonable Po...	character		0 1 0 0 0 0	0 1
AEOUTC	AE Outcome	character		0 0 0 0 1 1	0 1 2
AECODE	AE Code	character		10047332 10002546 100...	10047332 10002546 100699...
AEDTXT	AE Dictionary T...	character	name	Vertebral collapse Ankle...	Vertebral collapse Ankle oed...
PT_CODE	MedDRA Prefe...	numeric		10041541 10030124 100...	10041541 10030124 100646...
PT_NAME	MedDRA Prefe...	character		Spinal compression fract...	Spinal compression fracture...
SOC_CODE	MedDRA Syste...	numeric		10022117 10018065 1001...	10022117 10018065 100292...
SOC_NAME	MedDRA Syste...	character		Injury, poisoning and pro...	Injury, poisoning and procedu...
SOC_ABRV	MedDRA Syste...	character		Inj&P Genrl Genrl Nerv N...	Inj&P Genrl Genrl Nerv Neopl Skin...
HLGTCODE	MedDRA High...	numeric		10005942 10018073 100...	10005942 10018073 100697...
HLGTNAME	MedDRA High...	character		Bone and joint injuries Ge...	Bone and joint injuries Gener...
HLT_CODE	MedDRA High...	numeric		10041574 10030113 100...	10041574 10030113 1006979...
HLT_NAME	MedDRA High...	character		Spinal fractures and dislo...	Spinal fractures and dislocati...
LLT_CODE	MedDRA Lowe...	numeric		10047332 10002546 100...	10047332 10002546 100699...
LLT_NAME	MedDRA Lowe...	character		Vertebral collapse Ankle...	Vertebral collapse Ankle oed...
MEDDRA_V	MedDRA Version	character		14.0 14.0 14.0 14.0 14.0 1...	14
AE_SFL	AE start relativ...	character		D D D D A D	D A B
AE_EFL	AE end relative...	character		D D D D	D A B
AEDUR	AE duration	character		13 days 27 days 35 days...	13 days 27 days 35 days 84...
AEONSET	Onset after firs...	character		88 days 85 days 271 day...	88 days 85 days 271 days 6...
AESIG	Other significa...	character		0 0 0 0 0 0	0
AEDUR_D	AE duration (d...	numeric		13 27 35 84 NA NA	13 27 35 84 NA 32 38 4...
AESDYLTR	AE start - last...	numeric		-493 -331 -145 -217 22...	-493 -331 -145 -217 22...
POST28	AE occurred >...	character		0 0 0 0 0 0	0 1
AESDYTRT	AE start study...	numeric		88 85 271 6 244 32	88 85 271 6 244 32 5...
AEEDYTRT	AE end study d...	numeric		100 111 305 89 NA NA	100 111 305 89 NA 83 170 1...
AESDYRND	AE start study...	numeric	start_day	89 86 272 7 245 33	89 86 272 7 245 33 5...
AEEDYRND	AE end study d...	numeric	end_day	101 112 306 90 NA NA	101 112 306 90 NA 84 171 1...
AEEVNT01	Adverse Event	character	indic.num.1	1 1 1 1 1 1	1 0
AEEVNT02	Causally Relate...	character		0 1 0 0 0 0	0 1
AEEVNT05	Serious Advers...	character		0 0 1 0 1 0	0 1
AEEVNT06	SAE Not Leadi...	character		0 0 1 0 1 0	0 1
AEEVNT07	Causal SAE	character		0 0 0 0 0 0	0 1
AEEVNT10	AE Leading to...	character		0 0 0 0 0 0	0 1

Better option: tagging

```{r}
# Function to process tags
fn_standardize <- function(dat_raw, dat_meta){
  dat_raw = dat_raw[fn_applyFilters(dat_raw, dat_meta),]
  out = data.frame(patient    = dat_raw[meta$name[which(meta$tag == "patient")]],
                   start_day = dat_raw[meta$name[which(meta$tag == "start_day")]],
                   name      = dat_raw[meta$name[which(meta$tag == "name")]],
                   grade     = dat_raw[meta$name[which(meta$tag == "grade")]],
                   patient   = dat_raw[meta$name[which(meta$tag == "patient")]])
  if(any(meta$tag == "duration")) {
    out$end_day = out$start_day  +  start_day = dat_raw[meta$name[which(meta$tag == "duration")]]
    }else {
      out$end_day = dat_raw[meta$name[which(meta$tag == "end_day")]]
    }
  out
}

dat_amgen <- fn_standardize( dat_raw =  haven::read_sas("Prostat_Amgen_2006_155_aae.sas7bdat"),
                             dat_meta = read_csv(paste0("meta_","Prostat_Amgen_2006_155_aae.sas7bdat",".csv")))

dat_astraZe <- fn_standardize( dat_raw =  haven::read_sas("Prostat_AstraZe_2008_103_r_aevbb.sas7bdat"),
                               dat_meta = read_csv(paste0("meta_","Prostat_AstraZe_2008_103_r_aevbb.sas7bdat",".csv")))
# Combine
dat_ideal %>%
  full_join(dat_amgen) %>%
  full_join(dat_astraZe) |>
  dat_full
```

Tag superpowers: Generalization

Simple

Complex

Assign column names

Force formatting (string/num)

Create indicator variables

Create counts per patient

Long to wide form

Create filters

Don't leave data sitting on the table!

Better option: automatic tagging

Amgen Meta

meta_Prostat_Amgen_2006_155_aae.sas7bdat.csv

name	label	class	tag	head	unique_vals
STUDYID	Study Identifier	character		20050147 20050147 200501...	20050147
SUBJID	Subject Identifier for the Study	character	patient	001163 001163 001163 0011...	001163 00073...
GEOREG	Geographic Region	character		NORTH AMERICA NORTH A...	NORTH AMERI...
GEOREGCD	Geographic Region Code	character		NA NA NA NA NA NA	NA OT EU LA
TRTA	Actual Treatment Received	character		Controlled Arm Controlled Ar...	Controlled Arm
AGEY	Age at Enrollment in Years	numeric		86 86 86 86 86 86	86 77 72 87 8...
SEX	Sex	character		M M M M M M	M
RACE	Race	character		WHITE OR CAUCASIAN WHIT...	WHITE OR CA...
RACECD	Race Code	character		01 01 01 01 01 01	01 02 03 88 04
AEDECOD	Dictionary-Derived Term	character	name	Acquired phimosis Actinic ke...	Acquired phim...
AEBODSYS	Body System or Organ Class	character		Reproductive system and bre...	Reproductive s...
AELLT	Adverse Event Low Level Term	character		Paraphimosis Actinic keratosi...	Paraphimosis...
AEHLT	Adverse Event High Level Term	character		Penile disorders NEC (excl er...	Penile disorder...
AEHLGT	Adverse Event High Level Group Term	character		Penile and scrotal disorders (...	Penile and scr...
AESTDY	Study Day of Start of Adverse Event	numeric	start_day	471 581 187 NA 1050	471 581 187...
AEDURD	Duration of Event in Days	numeric	duration	5 1 1 NA NA 1	5 1 NA 11 1...
AECONT	Adverse Event Continuing	character			Y
AETOXGR	Standard Toxicity Grade	character	grade	02 02 02 02 02 01	02 01 04 03 05
AESER	Serious Event	character		N N N N Y N	N Y
AEREL	Causality	character		NOT RELATED NOT RELATED...	NOT RELATED...
AEACTLST	List of Action Code	character		88 88 88 01 04,88 88	88 01 04,88 0...
AESDTH	Results in Death	character		N N N N N N	N Y
AESLIFE	Is Life Threatening	character		N N N N N N	N Y

AstraZeneca Meta

meta_Prostat_AstraZe_2008_103_r_aevbb.sas7bdat.csv

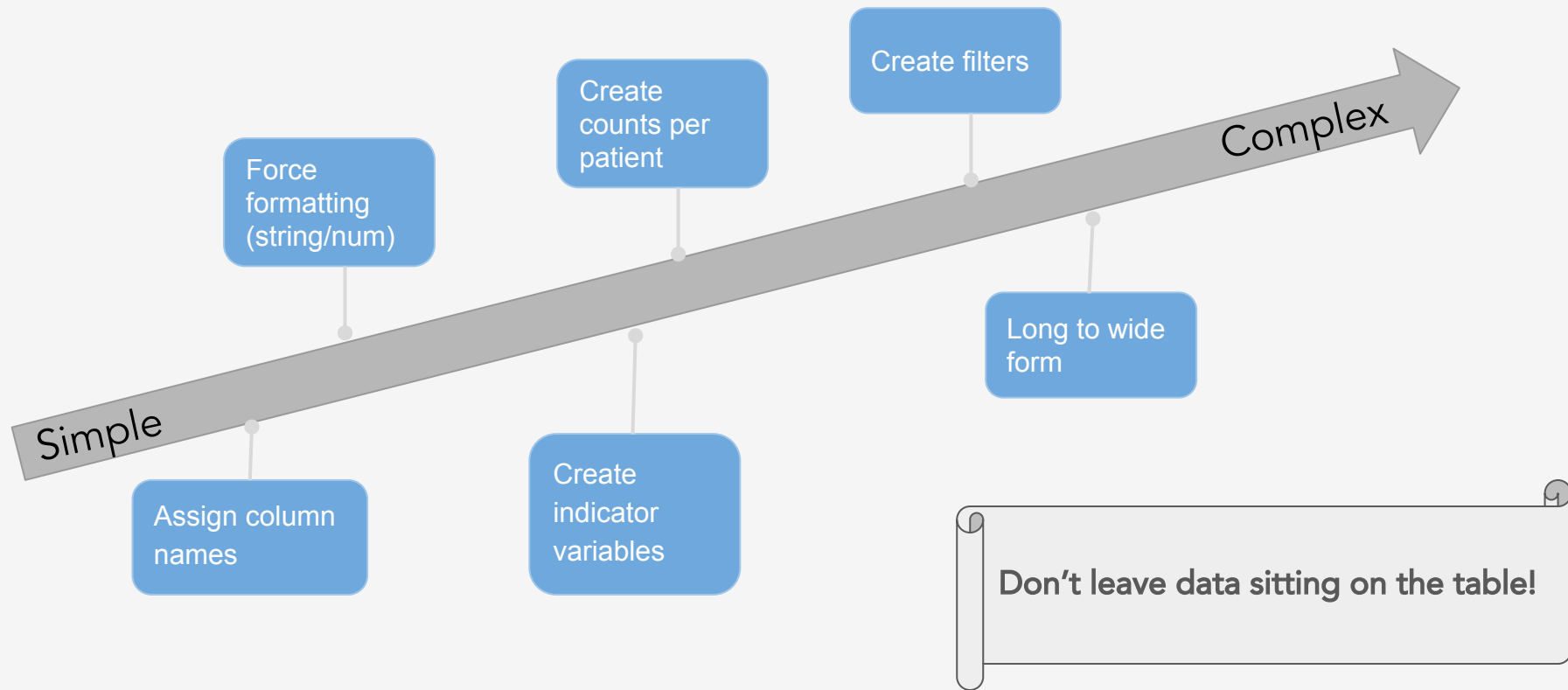| name	label	class	tag	head	unique_vals
AESER	Serious AE	character		0 0 1 0 1 0	0 1
SUBJ	Random patien...	numeric	patient	252 265 265 407 407 455	252 265 407 455 522 61 29...
AEACTIP	Action Taken, I...	character		0 0 0 0 0 0	0 3 2
CTCGMAX	Maximum CTC...	character	grade	2 1 2 2 3 1	2 1 3 5 4
AETREAT	Subject Recd T...	character		0 1 1 0 0 0	0 1
AECAUDIS	AE Caused Su...	character			0 1
AECAUS	Reasonable Po...	character		0 1 0 0 0 0	0 1
AEOUTC	AE Outcome	character		0 0 0 0 1 1	0 1 2
AECODE	AE Code	character		10047332 10002546 100...	10047332 10002546 100699...
AEDTXT	AE Dictionary T...	character	name	Vertebral collapse Ankle...	Vertebral collapse Ankle oed...
PT_CODE	MedDRA Prefe...	numeric		10041541 10030124 100...	10041541 10030124 100646...
PT_NAME	MedDRA Prefe...	character		Spinal compression fract...	Spinal compression fracture...
SOC_CODE	MedDRA Syste...	numeric		10022117 10018065 1001...	10022117 10018065 100292...
SOC_NAME	MedDRA Syste...	character		Injury, poisoning and pro...	Injury, poisoning and procedu...
SOC_ABRV	MedDRA Syste...	character		Inj&P Genrl Genrl Nerv N...	Inj&P Genrl Genrl Nerv Neopl Skin...
HLGTCODE	MedDRA High...	numeric		10005942 10018073 100...	10005942 10018073 100697...
HLGTNAME	MedDRA High...	character		Bone and joint injuries Ge...	Bone and joint injuries Gener...
HLT_CODE	MedDRA High...	numeric		10041574 10030113 100...	10041574 10030113 1006979...
HLT_NAME	MedDRA High...	character		Spinal fractures and dislo...	Spinal fractures and dislocati...
LLT_CODE	MedDRA Lowe...	numeric		10047332 10002546 100...	10047332 10002546 100699...
LLT_NAME	MedDRA Lowe...	character		Vertebral collapse Ankle...	Vertebral collapse Ankle oed...
MEDDRA_V	MedDRA Version	character		14.0 14.0 14.0 14.0 14.0 1...	14
AE_SFL	AE start relativ...	character		D D D D A D	D A B
AE_EFL	AE end relative...	character		D D D D	D A B
AEDUR	AE duration	character		13 days 27 days 35 days...	13 days 27 days 35 days 84...
AEONSET	Onset after firs...	character		88 days 85 days 271 day...	88 days 85 days 271 days 6...
AESIG	Other significa...	character		0 0 0 0 0 0	0
AEDUR_D	AE duration (d...	numeric		13 27 35 84 NA NA	13 27 35 84 NA 32 38 4...
AESDYLTR	AE start - last...	numeric		-493 -331 -145 -217 22...	-493 -331 -145 -217 22...
POST28	AE occurred >...	character		0 0 0 0 0 0	0 1
AESDYTRT	AE start study...	numeric		88 85 271 6 244 32	88 85 271 6 244 32 5...
AEEDYTRT	AE end study d...	numeric		100 111 305 89 NA NA	100 111 305 89 NA 83 170 1...
AESDYRND	AE start study...	numeric	start_day	89 86 272 7 245 33	89 86 272 7 245 33 5...
AEEDYRND	AE end study d...	numeric	end_day	101 112 306 90 NA NA	101 112 306 90 NA 84 171 1...
AEEVNT01	Adverse Event	character	indic.num.1	1 1 1 1 1 1	1 0
AEEVNT02	Causally Relate...	character		0 1 0 0 0 0	0 1
AEEVNT05	Serious Advers...	character		0 0 1 0 1 0	0 1
AEEVNT06	SAE Not Leadi...	character		0 0 1 0 1 0	0 1
AEEVNT07	Causal SAE	character		0 0 0 0 0 0	0 1
AEEVNT10	AE Leading to...	character		0 0 0 0 0 0	0 1

IMPORT → TIDY → [VISUALIZE / TRANSFORM / MODEL] → COMMUNICATE

EXPLORE

https://r4ds.had.co.nz/explore-intro.html

Thank you!

Stay in touch:

@ClaraOromendia

clara@projectronin.com

www.linkedin.com/in/clara-oromendia/