

Lab notebook

KBC

4/5/2021

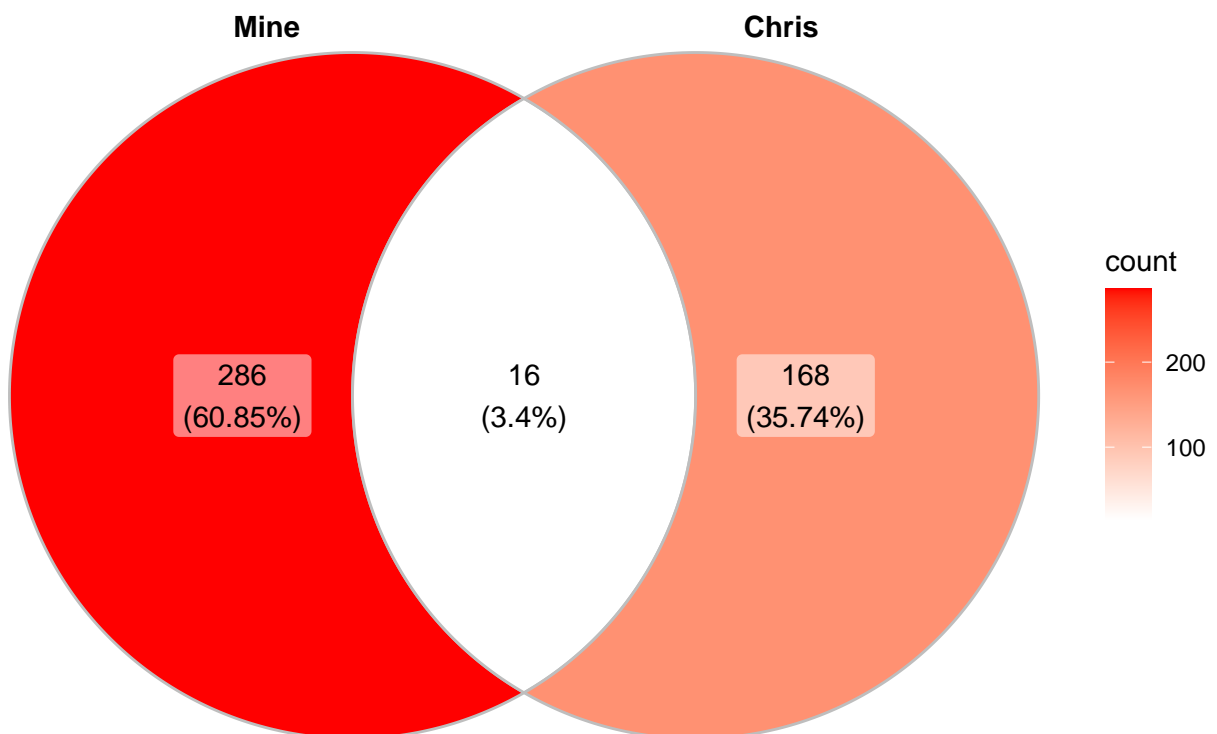
2021-04-05 report

Utility functions

Opposites already manually paired

Bill can explain where these came from better than I can.

Comparison to Chris et al.'s stuff



```
## [1] "Mine:"
```

```
## # A tibble: 302 x 2
##   term01      term02
##   <chr>      <chr>
## 1 HP:0030469 HP:0008275
## 2 HP:0100737 HP:0100736
## 3 HP:0025090 HP:0025129
## 4 HP:0031199 HP:0031197
## 5 HP:0011790 HP:0011791
## 6 HP:0012388 HP:0004469
## 7 HP:0100282 HP:0100281
## 8 HP:0012451 HP:0012450
## 9 HP:0032675 HP:0032676
## 10 HP:0011880 HP:0005520
## # ... with 292 more rows
```

```
## [1] "Chris's:"
```

```
## # A tibble: 184 x 2
##   term01      term02
##   <chr>      <chr>
## 1 HP:0000098 HP:0004322
## 2 HP:0000103 HP:0011037
## 3 HP:0000154 HP:0000160
## 4 HP:0000158 HP:0000171
## 5 HP:0000188 HP:0011341
## 6 HP:0000239 HP:0005486
## 7 HP:0000253 HP:0004481
## 8 HP:0000274 HP:0100729
## 9 HP:0000275 HP:0000283
## 10 HP:0000276 HP:0011219
## # ... with 174 more rows
```

OK, we know the intersection/complements, so let's see what Chris's opposites look like, exactly.

```
## Warning: package 'ontologyIndex' was built under R version 4.0.2
```

```
## [1] "FOUND BY BOTH ME AND CHRIS:"
```

```
## [1] "Decreased activity of mitochondrial respiratory chain      Increased activity of mitochondrial r
## [1] "Decreased circulating ACTH level      Increased circulating ACTH level"
## [1] "Decreased circulating antibody level      Increased circulating antibody level"
## [1] "Decreased HDL cholesterol concentration      Increased HDL cholesterol concentration"
## [1] "Decreased inflammatory response      Increased inflammatory response"
## [1] "Decreased lymphocyte apoptosis      Increased lymphocyte apoptosis"
## [1] "Decreased mean platelet volume      Increased mean platelet volume"
## [1] "Decreased mitochondrial number      Increased mitochondrial number"
## [1] "Decreased pineal volume      Increased pineal volume"
## [1] "Decreased size of nasopharyngeal adenoids      Increased size of nasopharyngeal adenoids"
## [1] "Decreased urinary sulfate      Increased urinary sulfate"
## [1] "Decreased urinary urate      Increased urinary urate"
## [1] "Large basal ganglia      Small basal ganglia"
## [1] "Large humeral epiphyses      Small humeral epiphyses"
## [1] "Large radial epiphyses      Small radial epiphyses"
## [1] "Large sella turcica      Small sella turcica"
```

```

## [1] "FOUND BY ME, NOT BY CHRIS:"

## [1] ""

## [1] "Abnormal dark-adapted electroretinogram      Abnormal light-adapted electroretinogram"
## [1] "Abnormal hard palate morphology      Abnormal soft palate morphology"
## [1] "Abnormal large intestinal mucosa morphology      Abnormal small intestinal mucosa morphology"
## [1] "Acellular urinary casts      Cellular urinary casts"
## [1] "Activating thyroid-stimulating hormone receptor defect      Inactivating thyroid-stimulating hormone receptor defect"
## [1] "Acute bronchitis      Chronic bronchitis"
## [1] "Acute colitis      Chronic colitis"
## [1] "Acute constipation      Chronic constipation"
## [1] "Acute cutaneous wound      Chronic cutaneous wound"
## [1] "Acute disseminated intravascular coagulation      Chronic disseminated intravascular coagulation"
## [1] "Acute hepatic failure      Chronic hepatic failure"
## [1] "Acute hepatitis      Chronic hepatitis"
## [1] "Acute leukemia      Chronic leukemia"
## [1] "Acute myelomonocytic leukemia      Chronic myelomonocytic leukemia"
## [1] "Acute otitis media      Chronic otitis media"
## [1] "Acute pancreatitis      Chronic pancreatitis"
## [1] "Acute respiratory acidosis      Chronic respiratory acidosis"
## [1] "Acute sinusitis      Chronic sinusitis"
## [1] "Acute tubulointerstitial nephritis      Chronic tubulointerstitial nephritis"
## [1] "Acute      Chronic"
## [1] "Adenocarcinoma of the large intestine      Adenocarcinoma of the small intestine"
## [1] "Affected      Unaffected"
## [1] "Agranulocytosis      Granulocytosis"
## [1] "Alobar holoprosencephaly      Lobar holoprosencephaly"
## [1] "Anhidrotic ectodermal dysplasia      Hidrotic ectodermal dysplasia"
## [1] "Asymmetric peripheral demyelination      Symmetric peripheral demyelination"
## [1] "Atonic seizure      Tonic seizure"
## [1] "Atypical absence seizure      Typical absence seizure"
## [1] "Atypical absence status epilepticus      Typical absence status epilepticus"
## [1] "Cleft hard palate      Cleft soft palate"
## [1] "Closed comedo      Open comedo"
## [1] "Communicating hydrocephalus      Noncommunicating hydrocephalus"
## [1] "Complete atrioventricular canal defect      Partial atrioventricular canal defect"
## [1] "Complete breech presentation      Incomplete breech presentation"
## [1] "Complete congenital stationary night blindness      Incomplete congenital stationary night blindness"
## [1] "Complete diaphragmatic absence of pericardium      Partial diaphragmatic absence of pericardium"
## [1] "Complete duplication of the 1st metatarsal      Partial duplication of the 1st metatarsal"
## [1] "Complete duplication of the distal phalanges of the hand      Partial duplication of the distal phalanges of the hand"
## [1] "Complete duplication of the distal phalanx of the 2nd finger      Partial duplication of the distal phalanx of the 2nd finger"
## [1] "Complete duplication of the distal phalanx of the 2nd toe      Partial duplication of the distal phalanx of the 2nd toe"
## [1] "Complete duplication of the distal phalanx of the 3rd finger      Partial duplication of the distal phalanx of the 3rd finger"
## [1] "Complete duplication of the distal phalanx of the 3rd toe      Partial duplication of the distal phalanx of the 3rd toe"
## [1] "Complete duplication of the distal phalanx of the 4th finger      Partial duplication of the distal phalanx of the 4th finger"
## [1] "Complete duplication of the distal phalanx of the 4th toe      Partial duplication of the distal phalanx of the 4th toe"
## [1] "Complete duplication of the distal phalanx of the 5th finger      Partial duplication of the distal phalanx of the 5th finger"
## [1] "Complete duplication of the distal phalanx of the 5th toe      Partial duplication of the distal phalanx of the 5th toe"
## [1] "Complete duplication of the distal phalanx of the hallux      Partial duplication of the distal phalanx of the hallux"
## [1] "Complete duplication of the middle phalanges of the hand      Partial duplication of the middle phalanges of the hand"
## [1] "Complete duplication of the middle phalanx of the 2nd finger      Partial duplication of the middle phalanx of the 2nd finger"
## [1] "Complete duplication of the middle phalanx of the 2nd toe      Partial duplication of the middle phalanx of the 2nd toe"

```

[1] "Complete duplication of the middle phalanx of the 3rd finger Partial duplication of the mid
 ## [1] "Complete duplication of the middle phalanx of the 3rd toe Partial duplication of the middle
 ## [1] "Complete duplication of the middle phalanx of the 4th finger Partial duplication of the mid
 ## [1] "Complete duplication of the middle phalanx of the 4th toe Partial duplication of the middle
 ## [1] "Complete duplication of the middle phalanx of the 5th finger Partial duplication of the mid
 ## [1] "Complete duplication of the middle phalanx of the 5th toe Partial duplication of the middle
 ## [1] "Complete duplication of the phalanges of the 2nd finger Partial duplication of the phalange
 ## [1] "Complete duplication of the phalanges of the 3rd finger Partial duplication of the phalange
 ## [1] "Complete duplication of the phalanges of the 4th finger Partial duplication of the phalange
 ## [1] "Complete duplication of the phalanges of the 5th finger Partial duplication of the phalange
 ## [1] "Complete duplication of the proximal phalanges of the hand Partial duplication of the proxim
 ## [1] "Complete duplication of the proximal phalanx of the 2nd finger Partial duplication of the p
 ## [1] "Complete duplication of the proximal phalanx of the 2nd toe Partial duplication of the prox
 ## [1] "Complete duplication of the proximal phalanx of the 3rd finger Partial duplication of the p
 ## [1] "Complete duplication of the proximal phalanx of the 3rd toe Partial duplication of the prox
 ## [1] "Complete duplication of the proximal phalanx of the 4th finger Partial duplication of the p
 ## [1] "Complete duplication of the proximal phalanx of the 4th toe Partial duplication of the prox
 ## [1] "Complete duplication of the proximal phalanx of the 5th finger Partial duplication of the p
 ## [1] "Complete duplication of the proximal phalanx of the 5th toe Partial duplication of the prox
 ## [1] "Complete duplication of the proximal phalanx of the hallux Partial duplication of the proxim
 ## [1] "Complete duplication of thumb phalanx Partial duplication of thumb phalanx"
 ## [1] "Complete left sided absence of pericardium Partial left sided absence of pericardium"
 ## [1] "Complete right sided absence of pericardium Partial right sided absence of pericardium"
 ## [1] "Complex febrile seizure Simple febrile seizure"
 ## [1] "Complex renal cyst Simple renal cyst"
 ## [1] "Conjugated hyperbilirubinemia Unconjugated hyperbilirubinemia"
 ## [1] "Cyclophoria Incyclophoria"
 ## [1] "Cyclotropia Incyclotropia"
 ## [1] "Decreased adiponectin level Increased adiponectin level"
 ## [1] "Decreased adipose tissue Increased adipose tissue"
 ## [1] "Decreased antimullerian hormone level Increased antimullerian hormone level"
 ## [1] "Decreased Arden ratio of electrooculogram Increased Arden ratio of electrooculogram"
 ## [1] "Decreased basophil count Increased basophil count"
 ## [1] "Decreased biotinidase level Increased biotinidase level"
 ## [1] "Decreased blood drug concentration Increased blood drug concentration"
 ## [1] "Decreased body fat percentage Increased body fat percentage"
 ## [1] "Decreased body mass index Increased body mass index"
 ## [1] "Decreased body weight Increased body weight"
 ## [1] "Decreased circulating 18-hydroxycortisone level Increased circulating 18-hydroxycortisone l
 ## [1] "Decreased circulating A-type natriuretic peptide level Increased circulating A-type natriur
 ## [1] "Decreased circulating androgen level Increased circulating androgen level"
 ## [1] "Decreased circulating beta-2-microglobulin level Increased circulating beta-2-microglobulin
 ## [1] "Decreased circulating beta-C-terminal telopeptide level Increased circulating beta-C-termin
 ## [1] "Decreased circulating ceruloplasmin concentration Increased circulating ceruloplasmin conce
 ## [1] "Decreased circulating chylomicron concentration Increased circulating chylomicron concentra
 ## [1] "Decreased circulating copper concentration Increased circulating copper concentration"
 ## [1] "Decreased circulating corticosterone level Increased circulating corticosterone level"
 ## [1] "Decreased circulating cortisol level Increased circulating cortisol level"
 ## [1] "Decreased circulating free fatty acid level Increased circulating free fatty acid level"
 ## [1] "Decreased circulating free T3 Increased circulating free T3"
 ## [1] "Decreased circulating free T4 level Increased circulating free T4 level"
 ## [1] "Decreased circulating globulin level Increased circulating globulin level"
 ## [1] "Decreased circulating gonadotropin level Increased circulating gonadotropin level"
 ## [1] "Decreased circulating heparan sulfate level Increased circulating heparan sulfate level"

[1] "Decreased circulating IgA level Increased circulating IgA level"
 ## [1] "Decreased circulating IgG level Increased circulating IgG level"
 ## [1] "Decreased circulating IgG1 level Increased circulating IgG1 level"
 ## [1] "Decreased circulating IgG2 level Increased circulating IgG2 level"
 ## [1] "Decreased circulating IgG3 level Increased circulating IgG3 level"
 ## [1] "Decreased circulating IgG4 level Increased circulating IgG4 level"
 ## [1] "Decreased circulating osteocalcin level Increased circulating osteocalcin level"
 ## [1] "Decreased circulating progesterone Increased circulating progesterone"
 ## [1] "Decreased circulating purine concentration Increased circulating purine concentration"
 ## [1] "Decreased circulating renin level Increased circulating renin level"
 ## [1] "Decreased circulating selenium concentration Increased circulating selenium concentration"
 ## [1] "Decreased circulating T4 level Increased circulating T4 level"
 ## [1] "Decreased complex N-glycan level Increased complex N-glycan level"
 ## [1] "Decreased core 1 O-glycan level Increased core 1 O-glycan level"
 ## [1] "Decreased corneal thickness Increased corneal thickness"
 ## [1] "Decreased CSF alanine concentration Increased CSF alanine concentration"
 ## [1] "Decreased CSF albumin concentration Increased CSF albumin concentration"
 ## [1] "Decreased CSF arginine concentration Increased CSF arginine concentration"
 ## [1] "Decreased CSF glutamate concentration Increased CSF glutamate concentration"
 ## [1] "Decreased CSF glutamine concentration Increased CSF glutamine concentration"
 ## [1] "Decreased CSF histidine concentration Increased CSF histidine concentration"
 ## [1] "Decreased CSF isoleucine concentration Increased CSF isoleucine concentration"
 ## [1] "Decreased CSF leucine concentration Increased CSF leucine concentration"
 ## [1] "Decreased CSF lysine concentration Increased CSF lysine concentration"
 ## [1] "Decreased CSF phenylalanine concentration Increased CSF phenylalanine concentration"
 ## [1] "Decreased CSF protein Increased CSF protein"
 ## [1] "Decreased CSF serine concentration Increased CSF serine concentration"
 ## [1] "Decreased CSF threonine concentration Increased CSF threonine concentration"
 ## [1] "Decreased CSF tyrosine concentration Increased CSF tyrosine concentration"
 ## [1] "Decreased CSF valine concentration Increased CSF valine concentration"
 ## [1] "Decreased CSF/serum albumin ratio Increased CSF/serum albumin ratio"
 ## [1] "Decreased cystatin C level Increased cystatin C level"
 ## [1] "Decreased DLCO Increased DLCO"
 ## [1] "Decreased female libido Increased female libido"
 ## [1] "Decreased femoral torsion Increased femoral torsion"
 ## [1] "Decreased fetal movement Increased fetal movement"
 ## [1] "Decreased fibular diameter Increased fibular diameter"
 ## [1] "Decreased fucosylation of N-linked protein glycosylation Increased fucosylation of N-linked
 ## [1] "Decreased fucosylation of O-linked protein glycosylation Increased fucosylation of O-linked
 ## [1] "Decreased glomerular filtration rate Increased glomerular filtration rate"
 ## [1] "Decreased glucagon level Increased glucagon level"
 ## [1] "Decreased glucose-6-phosphate dehydrogenase level in blood Increased glucose-6-phosphate del
 ## [1] "Decreased glucose-6-phosphate dehydrogenase level in dried blood spot Increased glucose-6-ph
 ## [1] "Decreased glucose-6-phosphate dehydrogenase level in leukocytes Increased glucose-6-phospha
 ## [1] "Decreased glucose-6-phosphate dehydrogenase level in red blood cells Increased glucose-6-ph
 ## [1] "Decreased glucose-6-phosphate dehydrogenase level in tissue Increased glucose-6-phosphate d
 ## [1] "Decreased HDL2a concentration Increased HDL2a concentration"
 ## [1] "Decreased HDL2b concentration Increased HDL2b concentration"
 ## [1] "Decreased HDL3a concentration Increased HDL3a concentration"
 ## [1] "Decreased HDL3b concentration Increased HDL3b concentration"
 ## [1] "Decreased HDL3c concentration Increased HDL3c concentration"
 ## [1] "Decreased head circumference Increased head circumference"
 ## [1] "Decreased heart rate variability Increased heart rate variability"
 ## [1] "Decreased helper T cell proportion Increased helper T cell proportion"

[1] "Decreased hemoglobin concentration Increased hemoglobin concentration"
 ## [1] "Decreased hepatic echogenicity Increased hepatic echogenicity"
 ## [1] "Decreased high-mannose N-glycan level Increased high-mannose N-glycan level"
 ## [1] "Decreased immunoglobulin level in body fluid Increased immunoglobulin level in body fluid"
 ## [1] "Decreased incisura length Increased incisura length"
 ## [1] "Decreased intestinal transit time Increased intestinal transit time"
 ## [1] "Decreased intracranial pressure Increased intracranial pressure"
 ## [1] "Decreased LDL cholesterol concentration Increased LDL cholesterol concentration"
 ## [1] "Decreased level of carnosine in blood Increased level of carnosine in blood"
 ## [1] "Decreased level of GABA in serum Increased level of GABA in serum"
 ## [1] "Decreased level of platelet-activating factor Increased level of platelet-activating factor"
 ## [1] "Decreased libido Increased libido"
 ## [1] "Decreased male libido Increased male libido"
 ## [1] "Decreased mannose-binding protein level Increased mannose-binding protein level"
 ## [1] "Decreased mannosylation of N-linked protein glycosylation Increased mannosylation of N-linked protein glycosylation"
 ## [1] "Decreased mean corpuscular hemoglobin concentration Increased mean corpuscular hemoglobin concentration"
 ## [1] "Decreased mean corpuscular volume Increased mean corpuscular volume"
 ## [1] "Decreased monosialylated core 1 O-glycan level Increased monosialylated core 1 O-glycan level"
 ## [1] "Decreased muscle glycogen content Increased muscle glycogen content"
 ## [1] "Decreased phosphoribosylpyrophosphate synthetase level Increased phosphoribosylpyrophosphate synthetase level"
 ## [1] "Decreased pituitary glycoprotein hormone alpha subunit level Increased pituitary glycoprotein hormone alpha subunit level"
 ## [1] "Decreased proportion of CD4+CD25+ regulatory T cells Increased proportion of CD4+CD25+ regulatory T cells"
 ## [1] "Decreased proportion of CD4-positive T cells Increased proportion of CD4-positive T cells"
 ## [1] "Decreased proportion of CD4-positive, alpha-beta memory T cells Increased proportion of CD4-positive, alpha-beta memory T cells"
 ## [1] "Decreased proportion of CD8-positive, alpha-beta memory T cells Increased proportion of CD8-positive, alpha-beta memory T cells"
 ## [1] "Decreased proportion of CD8-positive, alpha-beta TEMRA T cells Increased proportion of CD8-positive, alpha-beta TEMRA T cells"
 ## [1] "Decreased proportion of central memory CD4-positive, alpha-beta T cells Increased proportion of central memory CD4-positive, alpha-beta T cells"
 ## [1] "Decreased proportion of central memory CD8-positive, alpha-beta T cells Increased proportion of central memory CD8-positive, alpha-beta T cells"
 ## [1] "Decreased proportion of class-switched memory B cells Increased proportion of class-switched memory B cells"
 ## [1] "Decreased proportion of effector memory CD8-positive, alpha-beta T cells Increased proportion of effector memory CD8-positive, alpha-beta T cells"
 ## [1] "Decreased proportion of gamma-delta T cells Increased proportion of gamma-delta T cells"
 ## [1] "Decreased proportion of immature B cells Increased proportion of immature B cells"
 ## [1] "Decreased proportion of immature gamma-delta T cells Increased proportion of immature gamma-delta T cells"
 ## [1] "Decreased proportion of marginal zone B cells Increased proportion of marginal zone B cells"
 ## [1] "Decreased proportion of memory B cells Increased proportion of memory B cells"
 ## [1] "Decreased proportion of memory T cells Increased proportion of memory T cells"
 ## [1] "Decreased proportion of naive B cells Increased proportion of naive B cells"
 ## [1] "Decreased proportion of naive CD4 T cells Increased proportion of naive CD4 T cells"
 ## [1] "Decreased proportion of naive CD8 T cells Increased proportion of naive CD8 T cells"
 ## [1] "Decreased proportion of naive T cells Increased proportion of naive T cells"
 ## [1] "Decreased proportion of plasmablasts Increased proportion of plasmablasts"
 ## [1] "Decreased proportion of transitional B cells Increased proportion of transitional B cells"
 ## [1] "Decreased proportion of unswitched memory B cells Increased proportion of unswitched memory B cells"
 ## [1] "Decreased QRS voltage Increased QRS voltage"
 ## [1] "Decreased red blood cell count Increased red blood cell count"
 ## [1] "Decreased resting energy expenditure Increased resting energy expenditure"
 ## [1] "Decreased salivary cortisol level Increased salivary cortisol level"
 ## [1] "Decreased scrotal rugation Increased scrotal rugation"
 ## [1] "Decreased serum estradiol Increased serum estradiol"
 ## [1] "Decreased serum estriol Increased serum estriol"
 ## [1] "Decreased serum estrone Increased serum estrone"
 ## [1] "Decreased serum ferritin Increased serum ferritin"
 ## [1] "Decreased serum insulin-like growth factor 1 Increased serum insulin-like growth factor 1"
 ## [1] "Decreased serum iron Increased serum iron"

[1] "Decreased serum leptin Increased serum leptin"
 ## [1] "Decreased serum testosterone level Increased serum testosterone level"
 ## [1] "Decreased serum zinc Increased serum zinc"
 ## [1] "Decreased sialylated N-glycan level Increased sialylated N-glycan level"
 ## [1] "Decreased sialylation of N-linked protein glycosylation Increased sialylation of N-linked p
 ## [1] "Decreased size of the clitoris Increased size of the clitoris"
 ## [1] "Decreased skull ossification Increased skull ossification"
 ## [1] "Decreased superoxide dismutase level Increased superoxide dismutase level"
 ## [1] "Decreased T3/T4 ratio Increased T3/T4 ratio"
 ## [1] "Decreased total hemolytic complement activity Increased total hemolytic complement activity"
 ## [1] "Decreased total iron binding capacity Increased total iron binding capacity"
 ## [1] "Decreased uridine diphosphate glucose-4-epimerase level in plasma Increased uridine diphosph
 ## [1] "Decreased uridine diphosphate glucose-4-epimerase level in red blood cells Increased uridin
 ## [1] "Decreased urinary 1-methylhistidine Increased urinary 1-methylhistidine"
 ## [1] "Decreased urinary 3-methylhistidine Increased urinary 3-methylhistidine"
 ## [1] "Decreased urinary copper concentration Increased urinary copper concentration"
 ## [1] "Decreased urinary potassium Increased urinary potassium"
 ## [1] "Decreased urine alpha-ketoglutarate concentration Increased urine alpha-ketoglutarate conce
 ## [1] "Decreased urine urobilinogen Increased urine urobilinogen"
 ## [1] "Decreased VLDL cholesterol concentration Increased VLDL cholesterol concentration"
 ## [1] "Decreased waist to hip ratio Increased waist to hip ratio"
 ## [1] "Disproportionate short stature Proportionate short stature"
 ## [1] "Disproportionate tall stature Proportionate tall stature"
 ## [1] "Dysgerminoma Germinoma"
 ## [1] "Early chronotype Late chronotype"
 ## [1] "Early inspiratory crackles Late inspiratory crackles"
 ## [1] "Early spermatogenesis maturation arrest Late spermatogenesis maturation arrest"
 ## [1] "Fast-growing nails Slow-growing nails"
 ## [1] "Happy demeanor Unhappy demeanor"
 ## [1] "Heavy-chain paraproteinemia Light-chain paraproteinemia"
 ## [1] "High anterior hairline Low anterior hairline"
 ## [1] "High insertion of columella Low insertion of columella"
 ## [1] "High maternal serum alpha-fetoprotein Low maternal serum alpha-fetoprotein"
 ## [1] "High maternal serum chorionic gonadotropin Low maternal serum chorionic gonadotropin"
 ## [1] "High posterior hairline Low posterior hairline"
 ## [1] "High pulse pressure Low pulse pressure"
 ## [1] "High serum calcifediol Low serum calcifediol"
 ## [1] "High serum calcitriol Low serum calcitriol"
 ## [1] "High-frequency sensorineural hearing impairment Low-frequency sensorineural hearing impairm
 ## [1] "High-output congestive heart failure Low-output congestive heart failure"
 ## [1] "Hodgkin lymphoma Non-Hodgkin lymphoma"
 ## [1] "Ketotic hypoglycemia Nonketotic hypoglycemia"
 ## [1] "Large cell lung carcinoma Small cell lung carcinoma"
 ## [1] "Large earlobe Small earlobe"
 ## [1] "Large face Small face"
 ## [1] "Large for gestational age Small for gestational age"
 ## [1] "Large foramen magnum Small foramen magnum"
 ## [1] "Large forehead Small forehead"
 ## [1] "Large intestinal polyposis Small intestinal polyposis"
 ## [1] "Large placenta Small placenta"
 ## [1] "Large vessel vasculitis Small vessel vasculitis"
 ## [1] "Medial calcification of large arteries Medial calcification of small arteries"
 ## [1] "Medullary thyroid carcinoma Non-medullary thyroid carcinoma"
 ## [1] "Monoclonal elevation of IgG heavy chain Monoclonal elevation of IgG light chain"

```

## [1] "Neoplasm of the large intestine      Neoplasm of the small intestine"
## [1] "Non-obstructive azoospermia      Obstructive azoospermia"
## [1] "Non-restrictive ventricular septal defect      Restrictive ventricular septal defect"
## [1] "Non-secretory adrenocortical adenoma      Secretory adrenocortical adenoma"
## [1] "Non-small cell lung carcinoma      Small cell lung carcinoma"
## [1] "Nonopposable triphalangeal thumb      Opposable triphalangeal thumb"
## [1] "Nonproductive cough      Productive cough"
## [1] "Nonprogressive cerebellar ataxia      Progressive cerebellar ataxia"
## [1] "Nonprogressive encephalopathy      Progressive encephalopathy"
## [1] "Nonprogressive visual loss      Progressive visual loss"
## [1] "Nonprogressive      Progressive"
## [1] "obsolete Decreased proportion of CD4+ central memory cells      obsolete Increased proportion of
## [1] "Primary amenorrhea      Secondary amenorrhea"
## [1] "Primary Caesarian section      Secondary Caesarian section"
## [1] "Primary hyperaldosteronism      Secondary hyperaldosteronism"
## [1] "Primary hypercortisolism      Secondary hypercortisolism"
## [1] "Primary hyperparathyroidism      Secondary hyperparathyroidism"
## [1] "Secondary hyperparathyroidism      Tertiary hyperparathyroidism"
## [1] "Submucous cleft hard palate      Submucous cleft soft palate"
## [1] "Undetectable dark-adapted electroretinogram      Undetectable light-adapted electroretinogram"

## [1] "FOUND BY CHRIS, NOT BY ME:"

## [1] ""

## [1] "Tall stature      Short stature"
## [1] "Polyuria      Decreased urine output"
## [1] "Wide mouth      Narrow mouth"
## [1] "Macroglossia      Microglossia"
## [1] "Short upper lip      Long upper lip"
## [1] "Large fontanelles      Small fontanelle"
## [1] "Progressive microcephaly      Progressive macrocephaly"
## [1] "Small face      Large face"
## [1] "Narrow face      Broad face"
## [1] "Long face      Short face"
## [1] "Hypertelorism      Hypotelorism"
## [1] "Short philtrum      Long philtrum"
## [1] "Hypoplasia of the maxilla      Hyperplasia of the maxilla"
## [1] "Short chin      Tall chin"
## [1] "Broad forehead      Narrow forehead"
## [1] "Small forehead      Large forehead"
## [1] "Wide nose      Narrow nose"
## [1] "Short neck      Long neck"
## [1] "Microcornea      Megalocornea"
## [1] "Long eyelashes      Short eyelashes"
## [1] "Thick eyebrow      Thin eyebrow"
## [1] "Long palpebral fissure      Short palpebral fissure"
## [1] "Hypodontia      Increased number of teeth"
## [1] "Microdontia      Macrodontia"
## [1] "Hypoplasia of the thymus      Thymus hyperplasia"
## [1] "Hypothyroidism      Hyperthyroidism"
## [1] "Hypertension      Hypotension"
## [1] "Hypoparathyroidism      Hyperparathyroidism"

```


[1] "Increased circulating gonadotropin level Decreased circulating gonadotropin level"
 ## [1] "Hyperinsulinemia Hypoinsulinemia"
 ## [1] "Increased circulating renin level Decreased circulating renin level"
 ## [1] "Hyperaldosteronism Decreased circulating aldosterone level"
 ## [1] "Parathyroid hypoplasia Parathyroid hyperplasia"
 ## [1] "Long clavicles Short clavicles"
 ## [1] "Hypertrichosis obsolete Hypotrichosis"
 ## [1] "Broad palm Narrow palm"
 ## [1] "Large hands Small hand"
 ## [1] "Short umbilical cord Long umbilical cord"
 ## [1] "Hyporeflexia Hyperreflexia"
 ## [1] "Limitation of joint mobility Joint hypermobility"
 ## [1] "Onychogryposis Thin nail"
 ## [1] "Broad nail Narrow nail"
 ## [1] "Short toe Long toe"
 ## [1] "Long hallux Short hallux"
 ## [1] "Leukopenia Leukocytosis"
 ## [1] "Hypoglycemia Hyperglycemia"
 ## [1] "Hyperammonemia Hypoammonemia"
 ## [1] "Hypophosphatemia Hyperphosphatemia"
 ## [1] "Hyperuricemia Hypouricemia"
 ## [1] "Hypercalciuria Hypocalciuria"
 ## [1] "Hyperkalemia Hypokalemia"
 ## [1] "Hyperglycinemia Hypoglycinemia"
 ## [1] "Hypertriglyceridemia Hypotriglyceridemia"
 ## [1] "Increased CSF lactate Reduced CSF lactate"
 ## [1] "Pancreatic hypoplasia Pancreatic hyperplasia"
 ## [1] "Small foramen magnum Large foramen magnum"
 ## [1] "Thickened calvaria Thin calvarium"
 ## [1] "Lymphadenopathy Lymph node hypoplasia"
 ## [1] "Hypoventilation Hyperventilation"
 ## [1] "Decreased circulating total IgM Increased circulating IgM level"
 ## [1] "Hypocalcemia Hypercalcemia"
 ## [1] "Hyponatremia Hypernatremia"
 ## [1] "Hyperchloriduria Hypochloriduria"
 ## [1] "Hypomagnesemia Hypermagnesemia"
 ## [1] "Hypoalbuminemia Hyperalbuminemia"
 ## [1] "Hyperlipidemia Hypolipidemia"
 ## [1] "Increased urinary potassium Decreased urinary potassium"
 ## [1] "Hyperphosphaturia Hypophosphaturia"
 ## [1] "Hyperchloremia Hyperchloremia"
 ## [1] "Increased circulating cortisol level Decreased circulating cortisol level"
 ## [1] "Hypercholesterolemia Hypocholesterolemia"
 ## [1] "Elevated alkaline phosphatase Low alkaline phosphatase"
 ## [1] "Hypermethioninemia Hypomethioninemia"
 ## [1] "Increased circulating IgG level Decreased circulating IgG level"
 ## [1] "Elevated serum creatinine Decreased serum creatinine"
 ## [1] "Increased serum ferritin Decreased serum ferritin"
 ## [1] "CNS hypomyelination CNS hypermyelination"
 ## [1] "Short palm Long palm"
 ## [1] "Increased body weight Decreased body weight"
 ## [1] "Increased skull ossification Decreased skull ossification"
 ## [1] "Reduced bone mineral density Increased bone mineral density"
 ## [1] "Increased circulating purine concentration Decreased circulating purine concentration"

[1] "Biconcave vertebral bodies Biconvex vertebral bodies"
 ## [1] "Short middle phalanx of finger Long middle phalanx of finger"
 ## [1] "Thyroid hypoplasia Thyroid hyperplasia"
 ## [1] "Long proximal phalanx of finger Short proximal phalanx of finger"
 ## [1] "Small placenta Large placenta"
 ## [1] "Broad femoral neck Narrow femoral neck"
 ## [1] "Peripheral hypomyelination Peripheral hypermyelination"
 ## [1] "Generalized hyperpigmentation Generalized hypopigmentation"
 ## [1] "Long upper eyelashes Short upper eyelashes"
 ## [1] "Macular hypopigmentation Macular hyperpigmentation"
 ## [1] "Foveal hyperpigmentation Foveal hypopigmentation"
 ## [1] "Increased muscle glycogen content Decreased muscle glycogen content"
 ## [1] "Increased adipose tissue Decreased adipose tissue"
 ## [1] "Enlarged epiphysis of the distal phalanx of the 5th finger Small epiphysis of the distal phalanx of the 5th finger"
 ## [1] "Enlarged epiphysis of the proximal phalanx of the 5th finger Small epiphysis of the proximal phalanx of the 5th finger"
 ## [1] "Enlarged epiphysis of the middle phalanx of the 5th finger Small epiphysis of the middle phalanx of the 5th finger"
 ## [1] "Enlarged epiphysis of the middle phalanx of the 4th finger Small epiphysis of the middle phalanx of the 4th finger"
 ## [1] "Enlarged epiphysis of the distal phalanx of the 4th finger Small epiphysis of the distal phalanx of the 4th finger"
 ## [1] "Enlarged epiphysis of the proximal phalanx of the 4th finger Small epiphysis of the proximal phalanx of the 4th finger"
 ## [1] "Enlarged epiphysis of the middle phalanx of the 3rd finger Small epiphysis of the middle phalanx of the 3rd finger"
 ## [1] "Enlarged epiphysis of the distal phalanx of the 3rd finger Small epiphysis of the distal phalanx of the 3rd finger"
 ## [1] "Enlarged epiphysis of the proximal phalanx of the 3rd finger Small epiphysis of the proximal phalanx of the 3rd finger"
 ## [1] "Enlarged epiphyses of the 5th finger Small epiphyses of the 5th finger"
 ## [1] "Enlarged epiphyses of the 4th finger Small epiphyses of the 4th finger"
 ## [1] "Enlarged epiphyses of the 3rd finger Small epiphyses of the 3rd finger"
 ## [1] "Enlarged epiphyses of the 2nd finger Small epiphyses of the 2nd finger"
 ## [1] "Enlarged epiphysis of the distal phalanx of the 2nd finger Small epiphysis of the distal phalanx of the 2nd finger"
 ## [1] "Enlarged epiphysis of the middle phalanx of the 2nd finger Small epiphysis of the middle phalanx of the 2nd finger"
 ## [1] "Enlarged epiphysis of the proximal phalanx of the 2nd finger Small epiphysis of the proximal phalanx of the 2nd finger"
 ## [1] "Enlarged epiphysis of the proximal phalanx of the thumb Small epiphysis of the proximal phalanx of the thumb"
 ## [1] "Enlarged epiphysis of the distal phalanx of the thumb Small epiphysis of the distal phalanx of the thumb"
 ## [1] "Enlarged thumb epiphysis Small thumb epiphysis"
 ## [1] "Reduced number of teeth Increased number of teeth"
 ## [1] "Short distal phalanx of finger Long distal phalanx of finger"
 ## [1] "Enlarged epiphysis of the 1st metacarpal Small epiphysis of the 1st metacarpal"
 ## [1] "Short metacarpal Long metacarpals"
 ## [1] "Enlarged epiphyses of the hallux Small epiphyses of the hallux"
 ## [1] "Enlarged epiphysis of the proximal phalanx of the hallux Small epiphysis of the proximal phalanx of the hallux"
 ## [1] "Enlarged epiphysis of the distal phalanx of the hallux Small epiphysis of the distal phalanx of the hallux"
 ## [1] "Enlarged epiphysis of the 1st metatarsal Small epiphysis of the 1st metatarsal"
 ## [1] "Enlarged epiphyses of the toes Small epiphyses of the toes"
 ## [1] "Enlarged epiphyses of the phalanges of the hand Small epiphyses of the phalanges of the hand"
 ## [1] "Enlarged epiphyses of the distal phalanges of the hand Small epiphyses of the distal phalanges of the hand"
 ## [1] "Enlarged epiphyses of the middle phalanges of the hand Small epiphyses of the middle phalanges of the hand"
 ## [1] "Enlarged epiphyses of the proximal phalanges of the hand Small epiphyses of the proximal phalanges of the hand"
 ## [1] "Hyperpituitarism Hypopituitarism"
 ## [1] "Enlarged epiphyses Small epiphyses"
 ## [1] "Hypoplasia of the premaxilla Hyperplasia of the premaxilla"
 ## [1] "Broad uvula Narrow uvula"
 ## [1] "Long uvula Short uvula"
 ## [1] "Increased urinary copper concentration Decreased urinary copper concentration"
 ## [1] "Hyperlipoproteinemia Hypolipoproteinemia"
 ## [1] "Increased corneal thickness Decreased corneal thickness"
 ## [1] "Elevated circulating luteinizing hormone level Decreased circulating luteinizing hormone level"

```

## [1] "Erythroid hyperplasia      Erythroid hypoplasia"
## [1] "Granulocytic hyperplasia   Granulocytic hypoplasia"
## [1] "Monocytosis      Monocytopenia"
## [1] "Increased urine alpha-ketoglutarate concentration      Decreased urine alpha-ketoglutarate conce
## [1] "Hypocitraturia      Hypercitraturia"
## [1] "Enlarged pituitary gland      Small pituitary gland"
## [1] "Hyponatriuria      Hypernatriuria"
## [1] "Hypermagnesiuria      Hypomagnesiuria"
## [1] "Narrow jaw      Broad jaw"
## [1] "Increased serum testosterone level      Decreased serum testosterone level"
## [1] "Increased circulating androgen level      Decreased circulating androgen level"
## [1] "Chorioretinal hypopigmentation      Chorioretinal hyperpigmentation"
## [1] "Long lower eyelashes      Short lower eyelashes"
## [1] "Increased head circumference      Decreased head circumference"
## [1] "Increased size of the clitoris      Decreased size of the clitoris"
## [1] "Enlarged epiphyses of the 2nd toe      Small epiphyses of the 2nd toe"
## [1] "Enlarged epiphyses of the 3rd toe      Small epiphyses of the 3rd toe"
## [1] "Enlarged epiphyses of the 4th toe      Small epiphyses of the 4th toe"
## [1] "Enlarged epiphyses of the 5th toe      Small epiphyses of the 5th toe"
## [1] "Enlarged epiphysis of the distal phalanx of the 2nd toe      Small epiphysis of the distal phalan
## [1] "Enlarged epiphysis of the middle phalanx of the 2nd toe      Small epiphysis of the middle phalan
## [1] "Enlarged epiphysis of the proximal phalanx of the 2nd toe      Small epiphysis of the proximal p
## [1] "Enlarged epiphysis of the distal phalanx of the 3rd toe      Small epiphysis of the distal phalan
## [1] "Enlarged epiphysis of the middle phalanx of the 3rd toe      Small epiphysis of the middle phalan
## [1] "Enlarged epiphysis of the proximal phalanx of the 3rd toe      Small epiphysis of the proximal p
## [1] "Enlarged epiphysis of the distal phalanx of the 4th toe      Small epiphysis of the distal phalan
## [1] "Enlarged epiphysis of the middle phalanx of the 4th toe      Small epiphysis of the middle phalan
## [1] "Enlarged epiphysis of the proximal phalanx of the 4th toe      Small epiphysis of the proximal p
## [1] "Enlarged epiphysis of the distal phalanx of the 5th toe      Small epiphysis of the distal phalan
## [1] "Enlarged epiphysis of the middle phalanx of the 5th toe      Small epiphysis of the middle phalan
## [1] "Enlarged epiphysis of the proximal phalanx of the 5th toe      Small epiphysis of the proximal p
## [1] "Long ear      Short ear"

```

RO exploration

I extracted the terms from the current version of the RO, then got the absolute frequencies of their words:

```
/Users/kevincohen/Dropbox/Scripts-new/lexicalFrequency.pl experimental-outputs/ro.terms.txt
```

... which reveals imbalances in the use of words that are clear opposites, e.g.:

- during 11 before 4 after 2
- to 49 from 16 towards 1 *how about away?*
- directly 18 indirectly 8
- indirect 2 direct 2

... which suggests that I could be adding a bunch of opposite terms to the ontology.

So, I ran the current version of my antonym-finding script:

```
code/findAntonyms.pl resources/ontologies/ro.2021-03-08.obo.txt /wc -l
```

... which finds 43 pairs of opposites.

Next step: find overlap with Mike's resources/ontologies/predicates.txt.

For next weeks...

0. With Bill, reorganize repository and merge with TRANSLATOR's
1. Test cases in human-readable form
2. Handle substitutions (e.g. hypercalcemia vs. hypocalcemia)
3. Suffixes (e.g. hydrophilic vs. hydrophobic)
4. Word-internal, possibly (might produce a lot of false positives later when we do generation)

2021-04-15 report

Analysis of PATO opposites WRT the excluded middle

Motivation: this picks out a specific type of opposition: what Pustejovsky calls *polar*, meaning that there's a scale and the things at the two poles of the scale are opposites of each other.

How I assigned the excluded.middle value

0. If the middle is reasonably clearly excluded, I assigned the value *yes*. Examples: *acute/chronic*, *aerobic/anaerobic*. If that is not the case, then I put the "middle" value in the excluded.middle field. Example: for *phosphorylated/dephosphorylated*, I put the value *unphosphorylated* in the excluded.middle column.
1. Almost every pair that fits the pattern *increased/decreased x* has a mid-point or neutral point *normal x*.
2. Most pairs of the form *x/unx* (e.g. *responsive/unresponsive*) exclude the middle. Exception: *damaged/undamaged/repaired*.
3. Most pairs of the form *x/dex* (e.g. *phosphorylated/dephosphorylated*) have a neutral point *unx* (e.g. *unphosphorylated*).
4. Most pairs of the form *hypox/hyperx* where both members of the pair are single words have a single-word neutral point *normox*. For example, for the pair *hypotrophic/hypertrophic*, there is a neutral point *normotrophic*. If I had any question about the legitimacy of these, I checked Google Scholar to ensure that the *normox* word is used.
5. If I did not find such a word via Google Scholar, then I searched for the phrase *neither term01 nor term02*. If I found it, then I put the phrase in the field. (TODO: now that I think about it, if I didn't find the phrase, I didn't try again with the order *term02 term01*. Need to do that.)

```
# "current" is the ones that Bill found currently in use
in.both <- read.table("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-output")
in.mine.only <- read.table("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-output")
in.current.only <- read.table("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-output")

in.both$excluded.middle <- as.character(in.both$excluded.middle)
in.mine.only$excluded.middle <- as.character(in.mine.only$excluded.middle)
in.current.only$excluded.middle <- as.character(in.current.only$excluded.middle)
# in.both <- as_tibble(in.both)
# in.both <- mutate(in.both, found.by = "BOTH")
# in.mine.only <- as_tibble(in.mine.only)
# in.mine.only <- mutate(in.mine.only, found.by = "ME")
# in.current.only <- as_tibble(in.current.only)
# in.current.only <- mutate(in.current.only, found.by = "CURRENT")
```

```

# in.combined <- as_tibble(in.both, in.mine.only, in.current.only)

#ggplot(data = in.combined, mapping = aes(x = found.by, )) +
# geom_bar(stat = "identity")

in.counts <- c(nrow(in.both), nrow(in.mine.only), nrow(in.current.only))
#in.counts <- as_tibble(in.counts, counts = in.counts)
barplot(in.counts, names.arg = c("BOTH", "ME ONLY", "CURRENT ONLY"))

# I CAN'T GET THIS TO WORK...
#excluded.middle.counts <- c()
##count.both <- nrow(select(in.both, in.both$excluded.middle == "yes"))
##select(in.both, in.both$excluded.middle == "yes")
##count.both <- nrow(in.both$term01[in.both$term01 == "yes"])
##count.both <- in.both[which(in.both$excluded.middle == 'yes')]

#count.mine <- nrow(in.mine.only$term01[in.mine.only$term01 == "yes"])
#count.current <- nrow(in.current.only$term01[in.current.only$term01 == "yes"])
#excluded.middle.counts <- c(count.both, count.mine, count.current)

#barplot(excluded.middle.counts, names.arg = c("BOTH", "ME ONLY", "CURRENT ONLY"))

```

For next weeks...

1. Test cases in human-readable form
2. Handle substitutions (e.g. hypercalcemia vs. hypocalcemia)
3. Failing test case: *protein folding* versus *protein unfolding*
4. Failing test case: *name: oil gland decreased thickness* versus *oil gland increased thickness*
5. Suffixes (e.g. hydrophilic vs. hydrophobic)
6. Consistent naming scheme for experimental-results directory
7. Word-internal, possibly (might produce a lot of false positives later when we do generation)
8. *See email exchange with Bill*

2021-04-21 report

Now handling:

1. Now handling morphological substitutions, as opposed to additions. That means that where we used to get only pairs like abnormal, where the contrast is between presence of ab and absence of ab, we now also get hypercalcemia/hypocalcemia, where the contrast is not presence/absence, but rather between two things that are... Shit, I'm tired of trying to squeeze this into non-technical language. We used to only get prefix + free morpheme; now we are getting prefix + bound morpheme. Honi soit qui mal y pense.
2. Now handling suffixes. So, we now get pairs like hydrophilic/hydrophobic (thanks to Leslie Rapp for that one). Embarrassingly, I am not getting leukemia/leukopenia-bug being hunted.
3. Test cases in human-readable form
4. Failing test case: *protein folding* versus *p274rotein unfolding*
5. Failing test case: *name: oil gland decreased thickness* versus *oil gland increased thickness*

6. Consistent naming scheme for experimental-results directory
7. *See email exchange with Bill*
8. Word-internal, possibly (might produce a lot of false positives later when we do generation)

2021-05-05 report

1. Took a week of vacation
2. Took a sick day
3. Generated the outputs for all of the CRAFT ontologies, plus HPO, MPO, and PATO

2021-05-26

```
# only need to do once
#install.packages("entropy")
#library(entropy)

mi.calcs <- read.table("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-outp

get.rid.of.commas <- function(input.vector) {
  output.vector <- gsub(",", "", input.vector)
  return(output.vector)
}

# preprocessing--some things need to be integers, others factors
mi.calcs$opposite <- factor(mi.calcs$opposite)
mi.calcs$x.count <- as.integer(get.rid.of.commas(mi.calcs$x.count))
mi.calcs$y.count <- as.integer(get.rid.of.commas(mi.calcs$y.count))
mi.calcs$y.minus.x.count <- as.integer(get.rid.of.commas(mi.calcs$y.minus.x.count))
mi.calcs$x.minus.y.count <- as.integer(get.rid.of.commas(mi.calcs$x.minus.y.count))
mi.calcs$xy.count <- as.integer(get.rid.of.commas(mi.calcs$xy.count))
as_tibble(mi.calcs)

## # A tibble: 20 x 11
##   opposite x      x.count x.minus.y x.minus.y.count y      y.count y.minus.x
##   <fct>   <chr>   <int> <chr>          <int> <chr>   <int> <chr>
## 1 1      hot    192120 hot -cold      15658 cold   190861 -hot cold
## 2 1      leuk~  342485 leukemia~    5133 leuk~  50071 -leukemi~
## 3 0      hot    192120 hot -warm      4094 warm   37566 -hot warm
## 4 1      light  941650 light -d~    48407 dark   93314 -light d~
## 5 1      light  941650 light -h~    28783 heavy  194594 -light h~
## 6 1      arte~ 1151512 arterial~ 183230 veno~  459314 -arteria~
## 7 1      depr~  539864 depressi~   6381 mania  11677 -depress~
## 8 1      depr~  539864 depressi~   4779 prom~  174446 -depress~
## 9 1      dise~  7211377 disease ~ 1525665 heal~ 5288004 -disease~
## 10 1     abno~ 1367904 abnormal~ 258240 norm~ 2051560 -abnorma~
## 11 1     abse~  156430 absent ~    54648 pres~ 4536527 -absent ~
## 12 0     read~   57263 reader --     845 writ~   7834 -reader ~
## 13 0     read~   57263 reader --     415 eras~   4388 -reader ~
## 14 0     writ~    7834 writer --     476 eras~   4388 -writer ~
## 15 0     alive  55686 alive -l~    1402 live   186225 -alive l~
```

```
## 16 0      ad      1650424 ad -dead      57794 dead      968558 -ad dead
## 17 0      form~ 7372304 formatio~      252945 info~ 1823664 -formati~
## 18 0      irate      26 irate -a~      975 ate      11330 -irate a~
## 19 0      cat      139788 cat -dog      13511 dog      364688 -cat dog
## 20 0      cold      190861 cold -wa~      10555 warm      37566 -cold wa~
## # ... with 3 more variables: y.minus.x.count <int>, xy <chr>, xy.count <int>
```

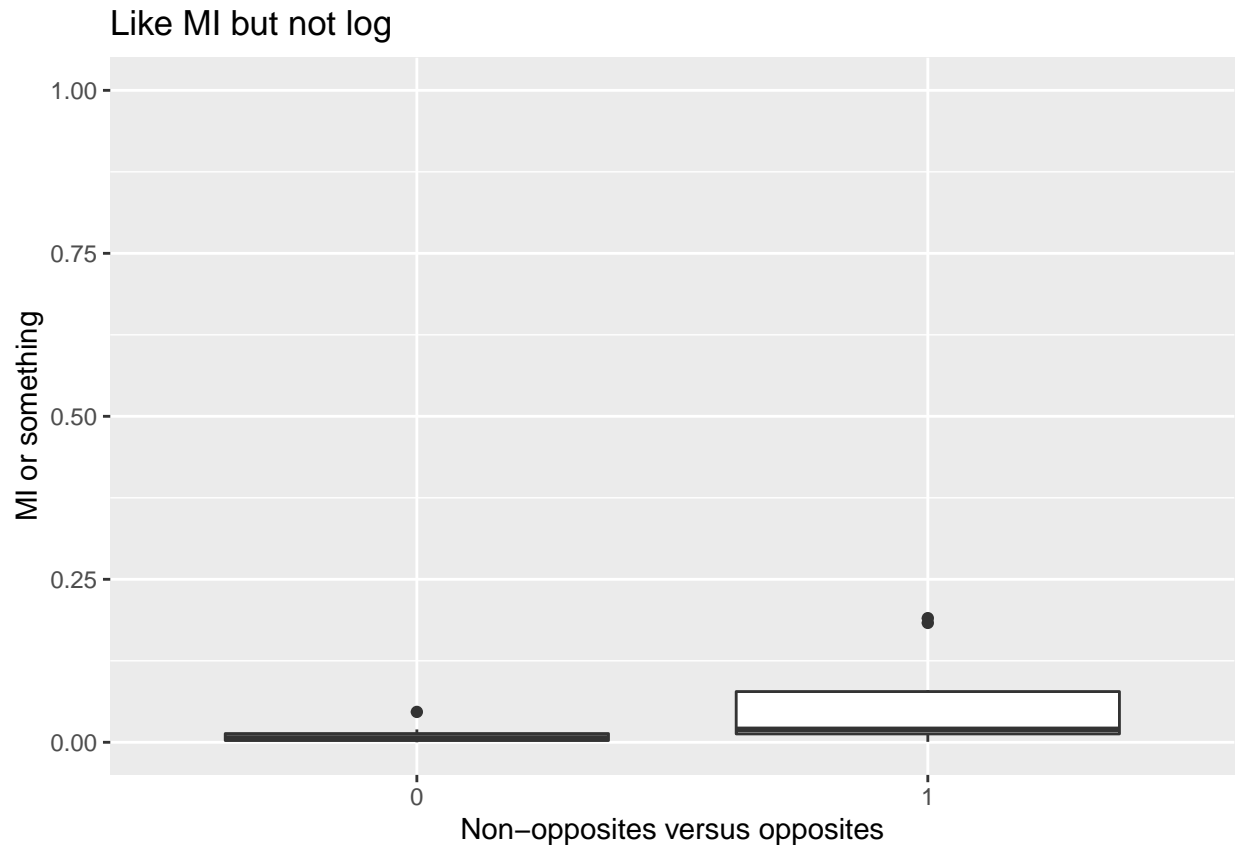
```
#mi.calcs <- mi.calcs %>% mutate(p.x = (x.count / (x.count + y.minus.x.count))) %>% mutate(p.y = (y.cou
#mutate(mi.or.something = p.xy / (p.x * p.y))
```

```
#mi.calcs <- mi.calcs %>% mutate(p.x = x.count / (x.not.y.count + y.not.x.count + x.and.y.count))
#mi.calcs <- mi.calcs %>% mutate(p.x = x.count)
# calculate p(x)
mi.calcs <- mi.calcs %>% mutate(p.x = x.count / (x.minus.y.count + y.minus.x.count + xy.count))
# calculate p(y)
mi.calcs <- mi.calcs %>% mutate(p.y = y.count / (x.minus.y.count + y.minus.x.count + xy.count))
# calculate p(x,y)
mi.calcs <- mi.calcs %>% mutate(p.xy = xy.count / (x.minus.y.count + y.minus.x.count + xy.count))
mi.calcs <- mi.calcs %>% mutate(mi.or.something = p.xy / (p.x * p.y))
```

```
#mi.calcs$mi.or.something <- log(mi.calcs$mi.or.something)
```

```
# Here I plot the values on a scale from 0.0 to 1.0
```

```
ggplot(data = mi.calcs, mapping = aes(x = opposite, y = mi.or.something)) +
  geom_boxplot() +
  ylim(0, 1.0) +
  #ylab("MI or something") +
  labs(x = "Non-opposites versus opposites", y = "MI or something", title = "Like MI but not log")
```



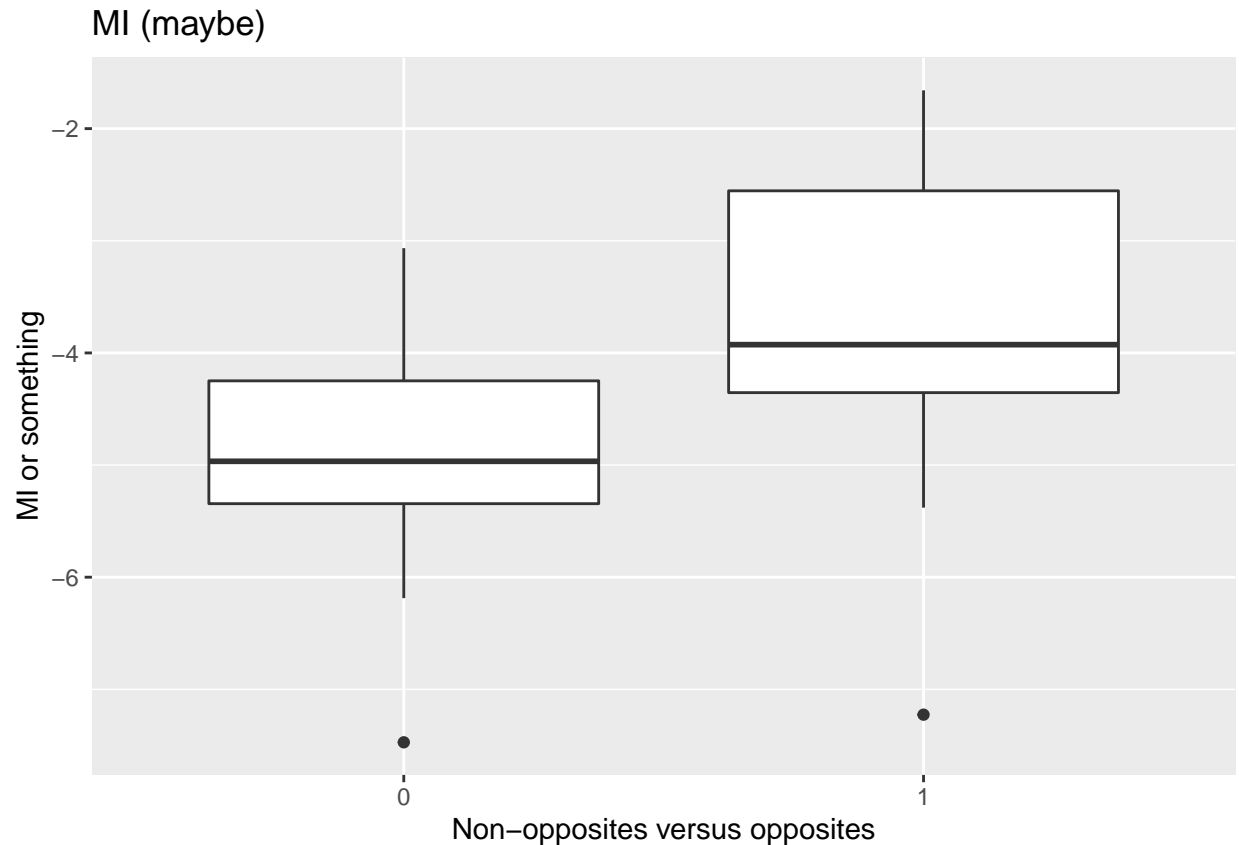
```
#names(c("Not opposites", "Opposites"))

shapiro.test(mi.calcs$mi.or.something)

##
##  Shapiro-Wilk normality test
##
## data:  mi.calcs$mi.or.something
## W = 0.63061, p-value = 6.216e-06

# Here I plot their logs
mi.calcs$mi.or.something <- log(mi.calcs$mi.or.something)
ggplot(data = mi.calcs, mapping = aes(x = opposite, y = mi.or.something)) +
  geom_boxplot() +
  ylim(0, 1.0) +
  ylab("MI or something") +
  labs(x = "Non-opposites versus opposites", y = "MI or something", title = "MI (maybe)")

## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
#names(c("Not opposites", "Opposites"))

head(mi.calcs)
```

##	opposite	x	x.count	x.minus.y	x.minus.y.count	y
## 1	1	hot	192120	hot -cold	15658	cold
## 2	1	leukemia	342485	leukemia -leukopenia	5133	leukopenia
## 3	0	hot	192120	hot -warm	4094	warm
## 4	1	light	941650	light -dark	48407	dark
## 5	1	light	941650	light -heavy	28783	heavy
## 6	1	arterial	1151512	arterial -venous	183230	venous
##	y.count	y.minus.x	y.minus.x.count	xy	xy.count	
## 1	190861	-hot cold	15658	hot cold	15657	
## 2	50071	-leukemia leukopenia	5133	leukemia leukopenia	5133	
## 3	37566	-hot warm	4094	hot warm	4094	
## 4	93314	-light dark	48407	light dark	48405	
## 5	194594	-light heavy	28783	light heavy	28784	
## 6	459314	-arterial venous	183230	arterial venous	183231	
##	p.x	p.y	p.xy	mi.or.something		
## 1	4.090009	4.0632065	0.3333191	-3.909175		
## 2	22.240730	3.2515748	0.3333333	-5.379677		
## 3	15.642404	3.0586224	0.3333333	-4.966562		
## 4	6.484344	0.6425743	0.3333242	-2.525758		
## 5	10.905038	2.2535495	0.3333411	-4.300321		
## 6	2.094835	0.8355858	0.3333345	-1.658461		

```
#mi.calcs$mi.or.something %>% gather(opposite)

#mi.calcs <- mi.calcs %>% select(opposite, mi.or.something)
#head(mi.calcs)

#wilcox.test(select(mi.calcs$opposite == "0"), select(mi.calcs$opposite == "1"))
```

Next step

Multi-word phrases, moving towards opposite sides of normal

HPO/MPO/MONDO terms with increase and decrease; then adjectival ones (especially hyper- and hypo-) (and generate more of those? Easy enough to do)

- Variability across those? Like, increased versus “increase in”, “elevation of”, etc.? Again, it’s generation...
- Additional affixes: over- and under-

2021-06-02

1. Observation: synonymy is impoverished in these ontologies. For example, *increased hemoglobin* is probably equivalent to *elevated hemoglobin*, but only the first is in the ontology.
2. So, the data would benefit from Dr. Funking.
3. Logical opposites don’t necessarily occur in these ontologies, and there might be good reasons for that. For example, *hypoxemia* is a clearly clinically relevant concept, but *hyperoxemia* might *not* be. This contrasts with *decreased hemoglobin affinity for oxygen* and *increased hemoglobin affinity for oxygen*, both of which *are* entirely clinically relevant.

Here are some numbers that support (1) and (3):

HPO contains:

- 459 non-obsolete terms with ‘increased’
- 370 non-obsolete terms with ‘decreased’
- 168 paired non-obsolete increased/decreased terms

This suggests that although perhaps there should *not* be more pairs of opposites, there certainly *could* be.

- 453 non-obsolete terms with ‘hyper’
- 671 non-obsolete terms with ‘hypo’
- 108 paired non-obsolete hypo/hyper terms

Again, this suggests that although perhaps there should *not* be more pairs of opposites, there certainly *could* be.

- 165 with ‘reduced’
- 153 with ‘elevated’
- 4 with ‘depressed’
- 33 with ‘high’
- 61 with ‘low’

The 153:33 ratio of *elevated* to *high* and 165:4:61 of *reduced/depressed/low* suggests that for recognition in text, Funkification would increase recall.

So, for any given ontology, here's what I did:

1. Grep out the terms with *increase*, *decrease*, *hyper*, or *hypo*.
2. Find the subset of those (increase/decrease and hyper/hypo) that do occur in addition to their logical opposite. (scripts: *increasedDecreasedOpposites.pl* and *hyperHypoOpposites.pl*)
3. For that subset, generate synonyms for both members of the pair. (script: *generatePairs.pl*)

Now let's go to a terminal...

2021-06-09

```
library("easyPubMed")

## Warning: package 'easyPubMed' was built under R version 4.0.2

files <- c("/Users/kevincohen/Downloads/TRANSLATOR opposites MI - PATO.tsv",
           "/Users/kevincohen/Downloads/TRANSLATOR opposites MI - HPO.tsv")

#sheet <- read.table("/Users/kevincohen/Downloads/TRANSLATOR opposites MI - PATO.tsv",
#                    header = TRUE, sep = "\t")

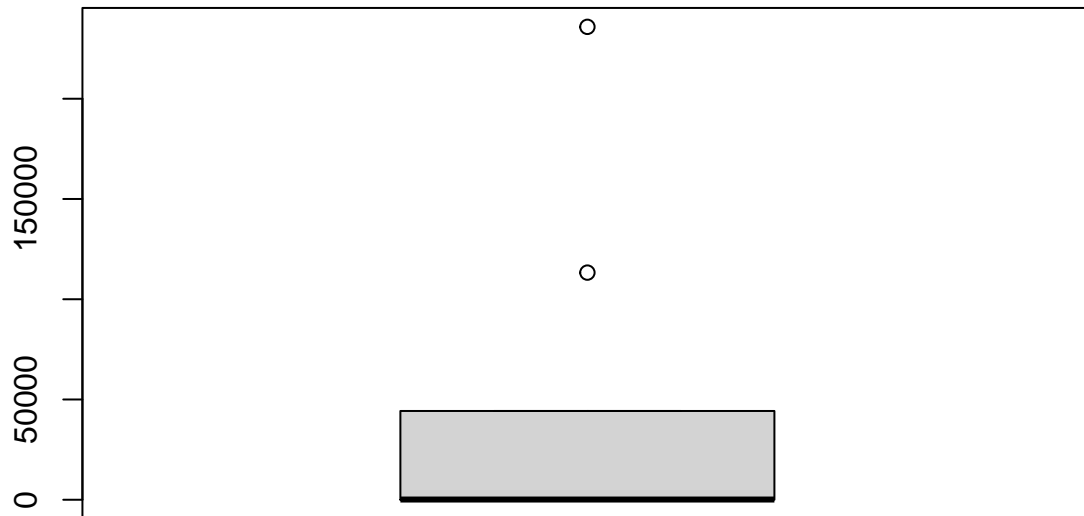
counts <- as.integer(c())

for (file_number in 1:length(files)) {
  sheet <- read.table(files[file_number], header = TRUE, sep = "\t")

  #for (i in 1:nrow(sheet)) {
  for (i in 1:5) {
    my_query <- paste("'", sheet[i, "term.01"], "'", " ", "'", sheet[i, "term.02"], "'", sep = "")
    print(my_query)
    my_entrez_id <- get_pubmed_ids(my_query)
    #print(my_entrez_id$Count)
    counts <- c(counts, as.integer(my_entrez_id$Count))
  } # loop through pairs of terms
} # close for-loop through list of files

## [1] "\"abaxial to\" \"axial to\""
## [1] "\"abnormal\" \"normal\""
## [1] "\"absent\" \"present\""
## [1] "\"active\" \"inactive\""
## [1] "\"acute\" \"chronic\""
## [1] "\"Embarrassing input: <leukopenia>\" \"\""
## [1] "\"Embarrassing input: <leukemia>\" \"\""
## [1] "\"abnormal dark-adapted electroretinogram\" \"abnormal light-adapted electroretinogram\""
## [1] "\"abnormal hard palate morphology\" \"abnormal soft palate morphology\""
## [1] "\"abnormal large intestinal mucosa morphology\" \"abnormal small intestinal mucosa morphology\""
```

```
#print(counts)
boxplot(counts)
```



For reproducibility

```
## R version 4.0.1 (2020-06-06)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] easyPubMed_2.13      ontologyIndex_2.7    ggVennDiagram_0.5.0
## [4] forcats_0.5.1        stringr_1.4.0        dplyr_1.0.4
## [7] purrr_0.3.4          readr_1.4.0          tidyr_1.1.1
## [10] tibble_3.0.1         ggplot2_3.3.2        tidyverse_1.3.0
```

```
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.4.6      lubridate_1.7.9    class_7.3-17
## [4] utf8_1.1.4        assertthat_0.2.1   digest_0.6.25
## [7] R6_2.4.1          cellranger_1.1.0   futile.options_1.0.1
## [10] backports_1.1.8    reprex_1.0.0       evaluate_0.14
## [13] e1071_1.7-3        highr_0.8           httr_1.4.1
## [16] pillar_1.4.4       rlang_0.4.10       readxl_1.3.1
## [19] VennDiagram_1.6.20 rstudioapi_0.11     rmarkdown_2.8
## [22] labeling_0.3        munsell_0.5.0       broom_0.7.4
## [25] compiler_4.0.1      modelr_0.1.8        xfun_0.23
## [28] pkgconfig_2.0.3     htmltools_0.5.1.1   tidyselect_1.1.0
## [31] fansi_0.4.1         crayon_1.3.4        dbplyr_2.1.0
## [34] withr_2.4.1         sf_0.9-8            grid_4.0.1
## [37] jsonlite_1.7.0      gtable_0.3.0        lifecycle_0.2.0
## [40] DBI_1.1.0           magrittr_1.5         formatR_1.8
## [43] units_0.7-1         scales_1.1.1        KernSmooth_2.23-17
## [46] cli_2.0.2           stringi_1.4.6        farver_2.0.3
## [49] fs_1.4.1            xml2_1.3.2           futile.logger_1.4.3
## [52] ellipsis_0.3.1      generics_0.0.2       vctrs_0.3.6
## [55] lambda.r_1.2.4       tools_4.0.1          glue_1.4.1
## [58] hms_1.0.0           yaml_2.2.1           colorspace_1.4-1
## [61] classInt_0.4-3      rvest_0.3.6          knitr_1.33
## [64] haven_2.3.1
```